Empirical Essays on Youths' Labour Markets and Education

Ștefania Simion
School of Economics and Finance
Queen Mary University of London

Submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy in Economics

April 2017

To Ștefana

Statement of originality

I, Ștefania Simion, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Details of collaboration and publications: The third chapter of the thesis is joint work with Professor Ghazala Azmat. My contribution to this paper included writing the literature review, cleaning and analysing the data in order to obtain the main results and working with Professor Azmat to interpret said results. The first chapter of the thesis has been presented at different conferences and previous versions of the work are available online. None of the chapters has been published up to date.

Signature: ______________________________

Date: ______________________________

Abstract

The first chapter assesses the impact of the cohort size on labour market outcomes. Using exogenous variation and micro-level data for France, the UK and the US, we study the effect of supply shocks measured at different ages on unemployment rates and wages during a cohort's life cycle. The results from an IV estimation show that the largest magnitude of the effects is found when the cohort size is measured at age 25. The impact of both wages and unemployment rates are temporary, however, both decreasing with time.

The second chapter analyses the effects of large inflows of foreign students on English undergraduates. Our results confirm previous findings that there is no overall effect, but we identify changes in the distribution of natives. We find that top performing English students are crowded in by foreign students. It is also mainly English-born males, natives who do not have English as their mother tongue and those of Asian ethnic origins that are crowded in by foreign students.

In chapter three, we aim to understand the short-term effects of changes in the level of the tuition fees charged by English universities on students' geographic mobility. Our results suggest that the increase in tuition fees in 2006/07 charged by English universities led students to enrol into universities that are closer to home, with a larger effect experienced by men and White students. Moreover, we find that students are less likely to move to universities located in rich areas.

Acknowledgements

First I would like to thank my thesis advisors Professor Marco Manacorda and Professor Ghazala Azmat. I am very grateful to Marco for opening my eyes to new ways of thinking, for continually offering me invaluable academic advice and for guiding me throughout my pursuit of the Ph.D. I am indebted to Ghazala for her help and academic guidance, for her trust and optimism in my abilities and my work from the very first day. The multitude of skills and knowledge I have acquired from and through them make me a truly fortunate student.

This thesis has benefited from valuable input received from members of Queen Mary University of London Economics' faculty and participants at various seminars and conference. A special thanks goes to Barbara Petrongolo, Erich Battistin, Francesco Fasani and Andrea Tesei for many fruitful discussions. I am also thankful to participants at the XI Jornadas de Economía Laboral, The Annual Scientific Conferences of Romanian Academic Economists from Abroad (ERMAS) and Queen Mary University of London Economics Reading Groups for their valuable comments.

The process of building further upon this thesis has been enhanced owing to the fruitful defense discussions. For this I want to extend my most sincere thanks to my examiners Maia Guell and Emla Fitzsimons.

I would also like to thank Katie Martin, Richard Garbitz and Pip Day for their help with the data application used in the analysis presented in the last two chapters. I would also like to thank Alaistair McCall for providing The Sunday Times Good University Guide data used in the second chapter. Financial support from the Postgraduate Research Fund at Queen Mary University of London and the Royal Economic Society is very gratefully acknowledged.

For the non-academic side of my thesis, I am thankful to my wonderful friends, Lorena, Carlos, Matias, Iustina and Felicia, who have been by my side through thick and thin.

Words do not do justice to the input my parents had into the completion of this thesis. Their continuous believe in my ability to finalise my studies, their unconditional love and tireless support have been the main drivers of my strive towards my goal.

And finally, I would like to thank Mihnea. I cannot imagine how my journey would have been without his love, patience, wisdom and understanding.

Contents

List of Figures

List of Tables

Chapter 1

Demographic bulges and labour market outcomes

1.1 Introduction

Differential patterns in fertility, mortality and immigration rates have contributed over time to changes in the population age structure of various countries. These changes in the age distribution of the population have implications for the labour market as workers in one age group are imperfect substitutes for workers in another age group. History shows that a demographic bulge can turn either into a demographic dividend or into a demographic burden, contributing even to a country's political instability. For instance, the large working-age population in East Asia between 1965 and 1990 played a major role in driving the East Asian economic miracle (Bloom and Williamson (1998)), while the large young population in the Arab region in the last decades played a major role in the recent Arab Spring uprise (Campante and Chor (2012)).

In economics it is a puzzle as to what direction the effects of demographic bulges on labour market outcomes go. Theory predicts that on the one hand a youth bulge can turn into a demographic dividend if firms respond to increases in youth relative supply by creating more jobs and consequently, spurring economic growth. On the other hand, a youth bulge can turn into a demographic burden if firms do not extend their capacities, as high capital investments may not be motivated in the long run. Consequently, the actual effect of cohort enlargement on labour market outcomes is an empirical question.

This paper uses detailed micro-level data for three large economies (France, the UK and the US) to causally estimate the effect of demographic bulges on cohort-specific labour market outcomes. In particular, through using an instrumental variable model we exploit variations in the timing and magnitude of cohort sizes within a country to identify the effect of the size of the cohort on unemployment rates and wages.

The relationship between age structure and labour market outcomes has received a great deal of attention in the literature in the last decades. The main focus has been on testing how youth bulges affect the performance in the labour market of younger workers,[1] by using the older cohorts as control groups. Production theory implies that an increase in the size of one age group relative to the other decreases relative productivity and consequently lowers relative factor prices (Ermisch (1988)). However, the underling assumption that prime-aged workers are not affected by demographic changes is problematic, as the theory does not predict whether the effect disappears as a cohort ages. Moreover, as only some papers in the literature control for confounding cohort and macroeconomic effects, mixed results have been found. While the literature on the impact of the cohort size on earnings identifies unanimously a negative effect in the short run ((Welch (1979), Freeman (1979), Card and Lemieux (2001), Fitzenberger and Kohn (2006), Brunello (2010))[2] there is no clear agreement regarding the persistency of the effect: Welch (1979) and Wright (1991) find temporary effects, but Berger (1985) concludes that the effect of the cohort size on earnings widens with experience. The strand of the literature which analyses the effect of the cohort sizes on unemployment rates finds that the impact is either positive (Bloom et al. (1987), Korenman and Neumark (2000), Jimeno and Rodriguez-Palenzuela (2002), Garloff et al. (2013), Newhouse and Wolff (2014)),[3] or negative (Shimer (2001), Skans (2005)).[4]

[1] Youths are defined as individuals aged 16-24 and the youth bulge is defined as a share of the population size of youngsters aged 16-24 relative to the population size of individuals ages 16-64.

[2] Welch (1979) and Freeman (1979) use the entry of the American baby-boomers on the labour markets to find that individuals belonging to larger cohorts face reductions in earnings rates, with a magnitude increasing with the level of schooling. Brunello (2010) shows that although the cohort size has a negative impact on wages in 11 European countries and larger for older workers, earnings tend to be more sensitive to changes in cohort sizes in countries with higher employment protection (i.e. Southern European countries). Moreover, Card and Lemieux (2001) provide evidence that the slowdown in the growth of college-educated labor for younger cohorts in the US, UK and Canada did not keep up with the steady skill bias in labour demand, triggering a rising return to college for younger men. Building on the work of Card and Lemieux (2001), Fitzenberger and Kohn (2006) show that Germans belonging to large cohorts experience wage reductions ranging from 8.8 to 12.2% and increase for the lower-skilled employees.

[3] Bloom et al. (1987) provide a review, while Korenman and Neumark (2000) use the older cohorts as a control group for younger workers to show that the cohort size has a positive effect on the unemployment rates of young workers. Jimeno and Rodriguez-Palenzuela (2002) conclude that in OECD countries fluctuations in cohort size positively associated with changes in relative youth unemployment rates. Newhouse and Wolff (2014) also find similar positive effects in developing countries, while Garloff et al. (2013) postulate that shrinking youth cohorts trigger lower overall unemployment rates in the economy.

[4] Shimer (2001) and Skans (2005) use variation at sub-national level and find that youth unemployment rates fall as cohort sizes increase in the US and Sweden, respectively. They argue that companies moved

This paper offers a possible reconciliation of the puzzle regarding these differential results. For this, we use data on three countries which have a very different labour market institutional setting: in the US the market is very liberalised, in France it is very rigid, and in the UK it is neither too liberalised, nor too rigid. Moreover, we investigate how the impact of the cohort size changes depending on the age at which the cohort size is measured. Our study is a comprehensive one as we consider various outcome variables, testing empirically the trade-off between wages and unemployment rates, by analysing the effect on each outcome through using the same estimation strategy and data set. And finally, we consider two mechanisms through which individuals can mitigate the potential burden of their generation size: education and migration.

We analyse the impact of the cohort size on labour market outcomes during a cohort's life cycle, rather than just at very young ages. The rich data available allows to follow large generations (the so called baby boomers) from early ages until they get close to their retirement ages. We consider individuals aged at least 25, age by which most people have finished education, in order to account for the endogenous investment in education which can be one of the mechanisms through which individuals manage their timing of entry into the labour market. Results presented in the empirical analysis show that individuals belonging to ballooning generations spend less time in education as a result of their larger cohort size and are more likely to have ended education with at most a high school or an equivalent degree. As for the persistency of the effects, the results show that the largest magnitude of the effects is registered when the cohort size is measured at age 25, decreasing as a cohort ages. The impact on both wages and the unemployment rates are temporary, with longer spans for wages.

Secondly, we test empirically the trade-off between wages and unemployment rates. Even though there is a large literature on the impact of the cohort size on wages and a large literature on the effect of the cohort size on unemployment rates, the different estimation strategies and data sets used across the two streams of literature make it difficult to understand whether those belonging to large generations are likely to face lower wages in the short-run and/or higher unemployment rates. This paper bridges these two streams of literature by investigating simultaneously the effect of the cohort size on unemployment rates and wages. Our findings suggest that the impact on unemployment rates are larger, compared to the one on wages. We show that the lower levels of education attained by baby boomers do not fully explain the negative effect on wages of the cohort size in the US.

to those areas with larger youth cohorts because there is a larger pool of well-educated graduates available to hire.

Given that the current population size of a given cohort is mainly a function of the number of live births, the number of deaths and the immigration waves, it may also affect the cohort's decision to migrate. Thus, we investigate if migration is one of the mechanism through which individuals can allay the potential effects of belonging to larger generations. In particular, we analyse how migration flows respond to larger cohort sizes through using only the sample of natives, accounting for the fact that immigrants and natives tend to be perceived as imperfect substitutes in the labour market (Manacorda et al. (2012)). Immigration rates are found to be lower due to larger cohort sizes among those individuals who have ended education with less than an undergraduate degree. In other words, the poor labour market prospects in the host countries discourage immigration flows for similarly educated individuals.

Finally, gender gaps in the labour market due to various factors like caring for young children, part-time work, occupational choice, differences in productivity or just preferences between men and women, have been largely documented (Manning and Petrongolo (2006), Azmat and Petrongolo (2014)). Thus, we present heterogenous effects by gender in order to understand if there is any differential effect of the cohort size for men and women. We find differential effects both across countries and gender. For instance, in France, a country with more rigid labour market institutions, the magnitude of the impact on women's unemployment rates is higher, compared to the UK and the US. Unemployment rates among women also show more persistent effects within countries, as women have a lower labour force attachment due to child rearing. Moreover, men face a larger penalty regarding the number of years spent in school compared to women, especially in France: a 10 percentage points increase in the cohort size triggers a decrease of around 0.45 years among men and of approximately 0.27 years among women. As for immigration rates, although all three countries are mainly immigrant receivers, it is among British women that the magnitudes are larger.

The remainder of the paper is organised as follows. Section 1.2 presents a simple theoretical model to rationalise the relationship between the age composition of labour supply and labour market outcomes across different age groups. Section 1.3 describes the data and defines the baby-boom generations in the three countries under analysis. Section 1.4 presents the estimation strategy, while section 1.5 discusses the results. Section 1.6 analyses the mechanisms and section 1.7 concludes the analysis.

1.2 Theoretical model

This section provides a theoretical framework for analysing the effect of the cohort size on labour market outcomes. The starting point is a simple supply-and-demand model of the labour market in which demographic bulges shift the labour supply curve outwards. In the model, aggregated output Y_t is produced using aggregated labour as an input:

$$Y_t = A_t N_t^{\alpha} \tag{1.1}$$

with A_t the total factor productivity and α the share of labour in the aggregated output. N_t captures the aggregated labour and it is assumed as a CES function of the quantities of labour provided by individuals of different ages:

$$N_t = \left[\sum_a (\theta_{at} N_{at})^{\frac{\sigma_a - 1}{\sigma_a}} \right]^{\frac{\sigma_a}{\sigma_a - 1}} \tag{1.2}$$

where N_{at} represents the number of individuals of age a employed at time t and θ_{at} is the productivity of workers of age a at time t. σ_a denotes the elasticity of substitution between workers of different ages and the lower it is, the more difficult it is to substitute workers of different ages. The disaggregation of the production function by many working groups implies the assumption that workers within the same age group are perfect substitutes, but workers across age groups are imperfect substitutes.

If labour markets are perfectly competitive and clear in every period, employed workers aged a at time t are paid a wage equal to their marginal product of labour. Thus, the demand for labour can be expressed as:

$$\ln\ N_{at} = \delta_t + (\sigma_a - 1) ln\ \theta_{at} - \sigma_a ln\ w_{at} \tag{1.3}$$

where $\delta_t = \sigma_a \left(ln\ \alpha A_t + (\alpha - 1 + \frac{1}{\sigma_a}) ln\ N_t \right)$ represents a common time-specific component shared by all age groups and w_{at} is the wage paid to employed workers aged a at time t.

As for the labour supply function, the approach of Card (2001) is followed and an upward sloping log-linear labour supply for each age group is used:

$$\frac{N_{at}}{P_{at}} = w_{at}^{\varepsilon} \tag{1.4}$$

with $\varepsilon > 0$ and P_{at} denoting the total number of people aged a as measured at time t. In other words, P_{at} represents the current cohort size for the group born in year $y-a$, as measured at time t.

Solving for the equilibrium, equations (1.3) and (1.4) lead to the below expression for the wage of employed workers:

$$ln\ w_{at} = \frac{1}{\varepsilon + \sigma_a}\Big(\delta_t + (\sigma_a - 1) ln\ \theta_{at} - ln\ P_{at}\Big) \tag{1.5}$$

This equation shows that wages of employed workers aged a at time t are determined by: a common time component (δ_t), an age and time-specific productivity shock (θ_{at}) and the current cohort size of that age group measured at time t (P_{at}).

Following Card (2001) we assume that the productivity component can be decomposed as:

$$ln\ \theta_{at} = \theta_a + \theta_t + \theta_{at}^{'} \tag{1.6}$$

where θ_a represents a common age effect, θ_t is a common time-effect and $\theta_{at}^{'}$ is an age and time-specific productivity term. Thus, equations (1.5) and (1.6) lead to a simplified expression of the wage as a function of age, a time component that is common across age groups and the population size of each age group:

$$ln\ w_{at} = m_t + m_a + \beta\ ln\ P_{at} + m_{at} \tag{1.7}$$

where m_t is a function of δ_t and θ_t and captures time fixed effects and m_a is a function of θ_a and captures age fixed effects. These fixed effects absorb any time or age components that might affect wages. Age-time specific unobserved productivity shocks are captured by m_{at} and depend on $\theta_{at}^{'}$, while β is a function of σ_a and ε.

Equation (1.7), which is derived from a simple supply-demand labour model, can be used as the theoretical base for an empirically testable expression of the impact of the current cohort size (captured by $ln\ P_{at}$) on wages. Given that the theory predicts that a demographic bulge could turn either into a dividend or a burden depending on the elasticity of the labour demand, the magnitude and sign of β, which describe the effect at interest, need to be tested empirically.

1.3 Data and descriptive statistics

In this section we first describe the data sources used in the analysis of the impact of the cohort size on labour market outcomes. We then proceed by defining the baby boom generations in each country and describing the main dependent variable considered in the analysis.

1.3.1 Data

We use data from three different source: the French Labour Force Survey (*L'enquête Emploie en continu*), the British Labour Force Survey and US March Current Population Survey to explore the variation in the cohort size across time over the following periods: 1990-2012 in France[5]; 1993-2013 in the UK[6] and 1994-2013 in the US[7].

These micro-level data sets, which contain detailed information, allow to aggregate the data at cohort-age-year cell in order to run cohort specific analysis. We define the cohort as the year of birth in order to gain more precision in the estimation. For instance, for the cohort born in UK in 1950 who were 57 years old in 2007, we use the 2007 British Labour Force Survey to calculate the cohort size as the weighted sum of all those aged 57 and born in 1950 who were surveyed in 2007 and this represents our 1950-57-2007 (cohort-age-time) observation cell. Our main sample are individuals born in the country of interest (i.e. natives) between 1940 and 1980. To account for the potential endogeneity of education - some people might delay entry on the labour market due to higher competition experienced by ballooning generations - we focus only on individuals aged at least 25 years old at survey time as by this age most individuals have finished at least their first-degree education. In the US and the UK the average age at graduating the first degree is 22, while in France it is 25 (OECD (2013)). Moreover, to account for early retirement we restrict the sample to individuals aged at most 60 years old at survey time. Thus, in total we have 745 observation cells for France, 692 for the UK and 663 for the USA.

[5] We use the annual surveys between 1990-2002, all four quarter surveys for 2003-2011 and the first quarter for 2012. The quarterly data and the one which includes all quarters, have different weights, such that the data will be representative for the population.

[6] The second quarter of the survey is used. Due to the small rotation in these labour force surveys, the use of one quarter along with the weights ensures the accuracy of the actual number of people and reduces the magnitude in measurement errors.

[7] We use the yearly March Current Population Survey.

For each cell we determine the current population size as the weighted number of individuals from that cell measured at survey time. We also calculate the average unemployment rate and the average level of weekly wages, our main outcome variables, at cell level.[8]

1.3.2 Descriptive statistics

As the empirical identification relies on variation in the cohort size in time, within a country, we proceed by documenting this variation in all three countries. The shifts in the demographic structure of a country's native population across time are a function of the number of live births, mortality and emigration rates registered by each cohort up to a specific time. That is, the population size for each cohort-age-time cell can be expressed as:

$$ln\ P_{act} = \gamma_0 + \gamma_c + \gamma_a + \gamma_t + \upsilon_{act} \tag{1.8}$$

where ln P_{act} denotes the natural logarithm of the current population size for cohort c of natives aged a in year of survey t ; γ_c, γ_a and γ_t are cohort, age and time effects, respectively; and υ_{act} is an idiosyncratic error term.

In equation (1.8) the cohort effect (γ_c) is indeed capturing demographic bulges, after the age and time effects have been been accounted for. However, due to perfect colinearity between age, cohort and time (age= time - cohort) we cannot identify one of the parameters. Given that the cohort effects are the variation in the size of the cohort that is unrelated to either time or age effects, we estimate the cohort effects through residuals of a regression in which the cohort size is predicted as a function of age and time:

$$ln\ P_{act} = \alpha_0 + \alpha_a + \alpha_t + \mu_{act} \tag{1.9}$$

[8] In all three countries, the employment status refers to one's main occupation at survey time. In the UK and France, the unemployed are identified by the International Labour Organization's unemployment definition as either out of work, but actively looking for a job or out of work and waiting to start a new job in the following two weeks. In the US, the unemployed are classified as those who did no work for pay or profit, did not have a job from which they were briefly absent, and answered yes to a question about whether they had been looking for work in the past four weeks; people who were temporarily laid off from a job were also classified as unemployed. In determining wages, the weighted mean of the gross weekly wages, expressed in PPP 1998 USA dollars, is calculated and all wage, salary and self-employed full-time workers are included (see Appendix A for details).

where α_a and α_t are age and time fixed effects, respectively; μ_{act} is an idiosyncratic error term. The regression is estimated with standard errors clustered at cohort level, as most of the variation in the cohort size is across cohorts, rather that within cohorts.

The residuals from equation (1.9) measure the variation in the current population by cohort and we plot them to investigate the variation in the cohort size. In figure (1.1) the shaded areas (which correspond to residuals above zero) identify the baby-boom generations. A first glimpse clearly brings into light the heterogeneity across countries in both the magnitudes and the timing of the ballooning generations. In France, the baby-boom (1946-1967) peaked in late 1940s and culminated with the Neuwirth Law which legalized contraception in late 1960s. The UK experienced two post-war baby booms, peaking twice: in 1947 and 1964. Apart from the high war participation rate of the British army forces in the World War II, the pill also played an important role in the timing of the baby boom. Although it was introduced in the UK in 1961, it was prescribed only to already mothers, and it took full effect only after its use was open to single women as well in the early 1970s. In the US, the baby-boom generation started as early as 1946 and reached its peak in late 1950s-early1960s. Yet, the introduction of the contraceptive pill in 1960 and its quick diffusion among married women as well as among single women in the late 1960s (Goldin and Katz (2002)) determined an earlier baby boost.

Figure (1.1) also provides evidence of the accuracy of these estimated residuals in identifying the baby boom generations. In particular, as the baby boom generations are determined by large numbers of live births for given generations and we observe individuals when they are at least 25, we further measure the cohort size through using the total number of live births between 1940-1980, rather than the current population size.[9] As the total number of live births of a given cohort is constant across time and ages we re-estimate the cohort effects as deviations around a linear cohort trend:

$$ln\ P_{ac0} = \gamma_0 + \gamma C + \eta_{ac0} \tag{1.10}$$

where $ln\ P_{ac0}$ is the natural logarithm of the number of live births; C is a cohort trend and η_{ac0} is an idiosyncratic error term. Figure (1.1) plots the residuals around the cohort trend from equation (1.10) as well. It can be easily observed that the two plotted residuals follow a similar trend and identify the same baby boom generations. Thus, the

[9] These figures are reported in various sources: UN Demographic yearbooks for the UK; the French National Institute of Statistics for France and the 2005 US National Vital Statistics Report, Vol. 53, No. 20, for the US.

evidence suggests that there is enough variation in the cohort size across time within the three countries to achieve precision.

Table (1.1) shows that there is also considerable variation across countries in the labour market outcomes. Panel A describes the average labour outcomes for all individuals. It seems that the unemployment rate is larger in France, while wages in France are lower on average compared to the other two countries. The last row of the panel shows that Americans spend on average 13.774 years in school, compared to 12.227 years spent in education among the British and only 9.744 years among the French.

As the paper explores heterogeneity across genders, panels B and C further describe the average labour outcomes for men and women, respectively. In France native men have lower and less volatile unemployment rates on average, compared to women.[10] This higher gender gap in unemployment rates between men and women in France, a country with a union coverage of up to 96% according to OECD data, is in line with the finding of Bertola et al. (2007) that countries with wider union coverage are associated with higher gender gaps in unemployment. Moreover, in the UK and the US, countries with lower union coverage, on average men have higher and more volatile unemployment rates than women, although the gap is positive and much lower in magnitude. On average, women in France face higher volatility in unemployment rates compared to British and American women.

Regarding wages, men tend to have higher weekly wages on average in all countries. The wage gap is higher in the UK and the U.S. which have a very decentralized collective wage bargaining and higher female labour force participation, compared to France. This fact is in line with the argument of Blau and Kahn (2003) that overall wage compression and low female supply relative to demand reduce a country's gender pay gap. Finally, as education is one of the mechanisms exploited in the paper, the last rows of each panels describe the average number of years spent in school by each gender group in each of the three countries. It seems that men spend more time in school on average than women, in all countries except for the US, where women tend to spend slightly more time in school.

1.4 Identification Strategy

In this section we discuss the empirical strategy used to estimate the effect of the cohort size on labour market outcomes. The starting point for the proposed estimation is the

[10] We refer to volatility as measured by the standard deviation.

relationship between wages and the cohort size developed through a simple supply-demand labour model in section (1.2) and presented in equation (1.7). We proceed by describing the instrumental variable strategy used to control for potential endogeneity of the cohort size.

1.4.1 Baseline model

As we have showed earlier in the paper, theory predicts a simple way to empirically test the effect of the cohort size on wages. Moreover, given the trade-off between wages and unemployment rates, we use equation (1.7) as the starting point for estimating the effect of the cohort size on all these labour market outcomes. By using the same equation to estimate the impact on all four outcomes we can understand which labour market outcome is more affected and how. We estimate the effect of the cohort size at age 25 on each outcome, allowing for linear heterogenous effects across age groups:

$$y_{act} = \beta_0 + \beta_1 \, ln \, P_{act} + \beta_2 \, ln \, P_{act} \, (a - 25) + d_a + d_t + \varepsilon_{act} \tag{1.11}$$

where y_{act} is a labour market outcome (either the unemployment rate or the natural logarithm of weekly wages) measured in year t for cohort c, aged a; $ln \, P_{act}$ is the natural logarithm of the cohort size for cohort c measured in year t and aged a; d_a and d_t are age and time fixed-effects, respectively, which control for any age and time specific characteristics that may affect the outcome variable; ε_{act} is an error term.

We cluster standard errors at cohort level and the regressions are run separately for each of the three countries - France, the UK and the US - and by gender. The main coefficient of interest, β_1, captures the impact of the cohort size at age 25 for cohort c, while β_2 captures changes in the effect of the cohort size as a cohort ages. The reason behind this distinction is that so far in the literature only the effect of the contemporaneous cohort size on labour market has been considered, without any focus on changes in the effect at different ages. By analysing the effect at the cohort size measured at different ages we can understand if the burden of belonging to a ballooning generation is different based on the age at which it is measured.

In estimating equation (1.11) we face the problem that the cohort size, measured by the current population size, is arguably endogenous to labour market conditions. That is, Ordinary Least Square (OLS) estimates of β_1 and β_2 would be biased. One major source of unobserved heterogeneity are unobserved shocks that affect the current labour market. For instance, the cohort size may be endogenous if migration flows respond to labour market conditions. In particular, poor labour market outcomes, like low wages or high unemployment rates, might lower the cohort sizes through discouraging immigration flows, encouraging emigration of individuals or even triggering poor life standards and consequently contributing to shorter spans of baby boomers. If the current population size is larger when labor markets are doing well, then the cohort size would be positively correlated with the level of wages, biasing the OLS estimate of β_1 and β_2 upward in the regression for wages. Similarly, the OLS estimates of β_1 and β_2 would display a downward bias in the estimation of the effect of the cohort size on the unemployment rate.

1.4.2 Identification issues

We use an instrumental variable strategy to address the issue of endogeneity of the cohort size. The ideal instrument is correlated with current population sizes and uncorrelated with current labour market outcomes.

We instrument the current cohort size by lagged birth rates, an approach used before in the literature of labour economics (see Korenman and Neumark (2000)). In particular, we account for migration decisions that could affect the current population size by restricting the sample only to natives and we instrument the cohort size of natives (denoted in equation (1.11) by P_{act}) by the number of live births registered for each birth cohort considered (denoted by P_{ac0}):

$$Z_{act} = P_{ac0} \tag{1.12}$$

The main assumption is that the number of live births for each cohort is exogenous to labour market conditions at least 25 years after birth. Moreover, as in the current paper we analyse heterogonous results separately by gender, it is worth noting that we use the total current population rather than separate measures by gender in order to measure the cohort size in the analysis. This is based on the fact that the sex ratio is relatively constant over time and that men and women do not compete in entirely different labour markets. Consequently, the instrument is also defined as the overall number of live births.

The relevance of the instrument rests on the notion that the current population size is related to the current cohort size. We can test empirically if the number of live births for each cohort is a good predictor of their cohort size later in life, by estimating the first stages. Table (1.2) reports the first stage estimates: in panel A we present the estimates of $ln\ Z_{act}$ when the outcome variable is $ln\ P_{act}$, while panel B presents the estimates of the interaction term $ln\ Z_{act}\ (a-25)$ when the outcome variable is $ln\ P_{act}\ (a-25)$. All coefficients are large in magnitude and statistically significant for all countries, implying relevant instruments.

Moreover, the large values of the F-statistics, which are clearly above the threshold value of 10, further suggest an accurate IV estimator. It should be noted that although the F-statistic for France is very large compared to the ones reported for the UK and the US, we do not believe this indicates a problem with our identification. Given that the cohort size of natives is a function of the number of live births, the mortality and the emigration rates, the instrument could have a higher predictive power of the endogeous variable in France compared to the UK and the US, if the variation in the mortality rate and emigration rates is lower in France. As all three countries are migrant receivers, the emigration rates have been relatively stable in all three countries in time. Thus, one potential explanation for the higher predictive power of the instrument in France could be a potentially lower variation in the mortality rate registered in France. Indeed, aggregated indicators suggest that in France the mortality rate has decreased sharply soon after the World War Two (affecting all the cohorts from our analysis) and stayed at a stable level afterwards. On the other hand, in the UK the mortality rate was quite large and stable until the early 1990s, decreasing afterwards, while in the US the mortality rate was also quite stable until the 1970s, decreasing afterwards.[11] In other words, while all French cohorts were exposed to lower mortality rates from early ages, in the UK and the US it was mainly younger cohorts that were exposed to lower mortality rates from early ages. This could explain why in France the live births predict much better the cohort size of natives.[12]

Regarding the exogeneity of the instruments, it seems plausible to assume that parents' fertility decisions are not caused by anticipated labour market conditions made at least 25 years ahead.

So far the discussion has assumed that $ln\ P_{act}$ is measured without error. This assumption is motivated by the fact that the data sets are representative for the population and

[11] Source: World Bank indicators, available at: http://data.worldbank.org/indicator

[12] One other reason, although we think this explains only a small part of the larger predictive power of the instrument in France versus the UK and the US, could be the fact that for France we use earlier survey years (from 1990 onwards) which means that the predictive power of the instrument is higher in these three years as cohorts are younger and thus more likely to be alive. However, restricting the data to the period 1993-2012 for France reduces the F statistic only slightly, from around 1150 to approximately 960.

each has weights which allow the extrapolation to the actual cohort size at country level. Thus, we believe this potential measurement error is negligible.

1.5 Results

In this section we examine how changes in cohort sizes affect labour market outcomes. We proceed by first analysing the effect as measured at age 25. We then test the sensitivity of the impact of the cohort size on labour to the state of the economy when a cohort turns 25. We finish by investigating the persistency of the effect.

1.5.1 Baseline results

As figure (1.2) shows there is a positive correlation between the cohort size and the unemployment rates, even after accounting for age and time effects, in all three countries. Table (1.3) presents the results of the estimated impact of the cohort size at age 25 on four different labour market outcomes: the unemployment rate and the logarithm of weekly wages. Each panel reports OLS and IV estimates of β_1 and β_2 from equation (1.11).

The first row of panel A presents estimates of the effect of the cohort size at age 25 on unemployment rates. Results point out that the magnitude of the effect is larger for women compared to men, across all three countries. The IV estimates indicate that in France, even though men do not face higher unemployment rates, among women, an increase of 10 percentage points in the cohort size at age 25 triggers an increase of around 0.011 percentage points in unemployment rates. In both the UK and the USA, a 10 percentage point increase in the size of the cohort leads to a increase of approximately 0.004 percentage points in men's unemployment rate. This magnitude is not negligible, given that the average unemployment rate among men is around 5.3% in both countries. Women face even higher impacts, of around 0.004 percentage points increase due to a 10 percentage point increase in the cohort size in the UK and the US. The much larger magnitude among French women is in line with the existing empirical literature on the negative impact of rigid labour market institutions on job creation and their association with higher levels of unemployment rates (Nickell (1997), Elmeskov and Scarpetta (1998), Nunziata (2002), Belot and van Ours (2004), Bassanini and Duval (2006), Azmat et al. (2006)). In other words, the more flexible American and British labour market helped keep the impact of increasing cohort sizes on unemployment rates among women lower than in France.

The second row of panel A reports estimates of β_2. In all countries and across gender the estimates are negative, implying that the effect of the cohort size decreases as a cohort ages. The impact is statistically significant for men in the UK and the US. However, among women, estimates for all three countries are statistically significant, implying a dissipating effect as a cohort ages.

The first row of panel B clearly shows that in two out of the three countries under analysis, i.e. in France and the US, for both men and women, increases in the cohort size at age 25 decrease weekly wages. The magnitudes are larger in the US compared to France, and among men compared to women in both countries. Specifically, a 10 percentage point increase in cohort sizes determines a decrease of 1.75 percentage points and of 3.97 percentage points in weekly wages among French and American men, respectively. On the other hand, among women the impact is much lower: a 10 percentage point increase in the cohort size triggers a decrease in weekly wages of 1.46 percentage points and of 2.40 percentage points in France and the US respectively. Yet, there is no effect of the cohort size at age 25 on weekly wages in the UK, for both men and women.

Regarding the persistency of the effect, the positive estimates of β_2 presented in the second row of panel B suggest that the effect disappears with age, although the estimates are statistically significant only for American men and women and British women.

Table (1.4) shows how the results change when the weighted regressions are run, with the weights equal to total number of men/women in each cohort-age-year cell. In all panels there are only small changes in the magnitude of the estimates, confirming the robustness of the baseline results.

The results presented so far indicate that an increase in the cohort triggers higher unemployment rates across all three countries, with large magnitudes of the effect experienced by women. As for wages, we find that in both the US and France men face larger impacts compared to women, but no statically significant effects are identified in the UK.

1.5.2 State of the economy at age 25

Given the large literature on the impact of the state of the economy at graduation on earnings (Oreopoulos et al. (2012), Kahn (2010), Genda et al. (2010)), the analysis is further extended to understand how sensitive the impact of the cohort size on labour market outcomes is to the inclusion of the state of the economy when a cohort turns 25 among the regressors.[13] The motivation behind the choice of the state of the economy at age 25 rather than at the time of graduating the highest level of education attained, is that the choice of the highest level of education attained is endogenous.[14] In particular, people might mitigate the timing of entering the labour market due to poor labour market conditions. That is, they may decide to extend their time in education if there are poor labour market prospects, delaying their entry into the labour market. The state of the economy at age 25 is proxied by the annual percentage growth rate of GDP.[15]

Table (1.5) presents both the OLS and the IV estimates of this adjusted regression. The format of the table is similar to table (1.3) and in addition it also reports the coefficient of the GDP measured at age 25. The estimates of the cohort size effect in all panels are comparable to the ones reported in table (1.3), while the estimates for the state of the economy at age 25 are mainly statistically insignificant. This leads to the conclusion that the impact of the cohort size is not affected by the state of the economy at age 25.

1.5.3 Persistency

So far in the analysis we have mainly focused on the effect of the cohort size on labour market outcomes at age 25, showing suggestive evidence that the effect dissipates with age. In this subsection we further extend the analysis to investigate how the effect changes at each age during a cohort's life time and at which age it disappears.

Figure (1.3) plots the estimates of $\beta_1 + \beta_2 * a$, where a is the difference between the age at which we measure the effect (i.e. an integer between 25 and 60) and 25, from equation (1.11) when the outcome variable is the unemployment rate. The plot shows that French men do not face statistically significant higher unemployment rates, independently of the age at which the cohort size is measured. However, among British men the cohort size measured at ages up until a cohort reaches their late 30s increases the cohort's

[13] The literature finds unanimously that a recession at high school or university graduation decreases earnings.

[14] Kahn (2010) also instruments for the state of the economy at college graduation time by the state of the economy when an individual turned the mode age of college graduation -22 years old- in order to measure what is the impact of the state of the economy on graduation time on earnings among college graduates.

[15] This is measured at market prices based on the 1990 Int.Geary-Khamis (GK) US dollar. A (GK) dollar, more commonly known as the international dollar, is a hypothetical unit of currency that has the same purchasing power parity that the U.S. dollar had in the United States at a given point in time. Source: The Maddison-Project, `http://www.ggdc.net/maddison/maddison-project/home.htm`, 2013 version. Thus, we re-estimate equation (1.11) by including this additional independent variable.

unemployment rates. Among men in the US, the impact of the cohort size is much shorter lived, having no statistically significant effect when the cohort size is measured after a cohort reaches their late 20s. What is of interest is that both in the UK and the US, the impact of the cohort size on men's unemployment rates turns negative and statistically significant if the cohort size is measured at ages higher than 50. We believe one explanation for this reverse trend could be due to delayed retirement, following increases in the retiring ages in the two countries.

Figure (1.3) further indicates that for women the effect on the unemployment rates is positive and statistically significant even when the cohort size is measured at later ages: in France it disappears when a cohort reaches their mid-50s, in the UK it dissipates during early 40s, while in the US it lasts until the cohort size is measured during its late 40s.

One potential explanation for the finding that the impact of the cohort size on unemployment rates is more persistent for women than men within countries could be the fact that women have a lower labour force attachment due to child rearing. As Azmat et al. (2006) show, we also find that for the demographic groups which are more likely to have lower labour force attachment (i.e. young ages at which women are more likely to have young children) the gender gap in the unemployment rates is higher. Moreover, we also confirm their findings that across countries the difference between the effect for men and women, is lower in the UK and the US, countries with higher levels of female labour force attachment compared to France.

Regarding the persistency of the effect on weekly wages for full-time workers, figure (1.4) shows that the impact is more persistent among men in both France and the US: it becomes statistically insignificant as a cohort reaches early 50s for men versus when a cohort reaches early 40s for women. For the UK, our findings suggest that the impact is not statistically significant when measured at any age for both men and women.

All in all, our findings indicate that the effects of the cohort size on the unemployment rates are more persistent for women in all three countries, although the magnitude of the effects drops as one ages. For men in the US and the UK the effect reverses sign at late ages. Moreover, the persistency of the effect of the cohort size on weekly wages is higher for men in the US and France, with no effect for either gender in the UK.

1.6 Mechanisms

This section analyses two potential mediating mechanisms which can explain our results: education and the migration. Members of large cohorts could stay in school for longer. Moreover, large cohorts of natives may discourage immigration flows if there are adverse economic effects of demographic bulges.

1.6.1 Education

Household investment in education is one of the mechanisms through which larger cohorts can manage their timing into the labour market. Specifically, declining wages due to larger cohort sizes could lead young people to stay for longer in school, postponing their entry into the labour market. Higher levels of education could also give them an edge into the competition on the labour market (for instance, higher returns to education or even higher chances to get a job for which one is overqualified). However, if tighter budgets prevent individuals from continuing their schooling (for instance, in the UK and the US the tuition fees are very high) or if larger cohorts compete for limited educational public resources (for example, heavily regulated places available in universities in the UK), baby boom generations may actually spend less time in schooling.

As the theory fails to predict how the size of the cohort affects educational attainment, this issue is further addressed empirically. For this, equation (1.11) is estimated to identify if baby boom generations spend more or less time in school. Firstly, panel A of table (1.6) presents IV estimates of β_1 and β_2 when the outcome variable is the average number of years spent in school.[16] The results are consistent across countries and gender, showing a statistically significant decrease in the number of years spent in education due to larger cohort sizes measured at age 25. For instance, in France a 10 percentage points increase in the cohort size triggers a decrease of around 0.45 years among men and of approximately 0.27 years among women. The lowest magnitudes are registered in the UK, where a 10 percentage increase in the cohort size leads to a

[16] It is worth noting that in the UK, the number of years spent in school is defined as age at which full-time education has been achieved minus five (the starting school age in the UK), while in France and the USA the average number of years spent in school is imputed based on the highest qualification declared. The reason for this differential definition between the three countries is that following section on immigration looks at heterogenous impacts by education and the complexity of the British educational system, tends to trigger differences between the highest qualification declared by natives and immigrants. In particular, while native British who classified their highest educational level attained as "other qualifications" category almost certainly have a low level of education- as all the major UK educational qualifications are covered by the alternative categories-, immigrants who are classified as being in the "other qualifications" category are more likely to have high levels of qualifications (Manacorda et al. (2012), Saleheen and Shadforth (2006)).

decrease of around 0.18 and 0.17 among men and women, respectively. In other words, baby boomers spent less time in education.

Secondly, panel B of table (1.6) offers further insights regarding the actual highest level of education attained by baby boom generations, through reporting estimates of the effect of the cohort size on the proportion of people in a given cohort with less than an undergraduate degree. It should be noted that the sample has been split into two categories: those with less than an undergraduate degree and those with at least an undergraduate degree.[17] According to the estimates presented in panel B, baby boomers from all countries are more likely to be either high school dropouts or to end education with a high school or some college degree in all three countries. The results obtained for the UK and the US are in line with the findings of Card and Lemieux (2001), who show that, in these two countries, people born between early 1950s and late 1960s were more likely to end education with a high school degree, compared to older generations (i.e. the baby boomers had a lower growth of the share of college graduates to high school graduates).

These empirical findings also support the theoretical model of Bound and Turner (2007) in which universities do not increase enrolments when the cohort sizes increase. The reason for this is that the higher education markets in all countries are dominated by public and non-profit production, and universities receive considerable subsidies from the state and private sources. Thus, consumers pay only a fraction of the costs faced by universities and the changes in demand due to higher cohort sizes are unlikely to be fully accommodated by tertiary education institutions without significant increases in non-tuition revenues.

The question to be raised at this point in the analysis is if the reduction in the number of years spent in school can be propagated on the labour market outcomes. In other words, is it the case that the reduction in wages due to larger cohort sizes in France and the US documented in section (1.5) is a result of the lower number of years spent in schooling? One way of checking if the negative impact of the cohort size on wages is triggered by the lower educational attainment of baby boomers or by both lower educational attainment and higher cohort sizes is to do a back-of-the-envelope calculation. This can be done by comparing the ratio of the estimate of the cohort size when the dependent variable is the logarithm of the wages (the estimate reported in the first row of Panel

[17] Just as in the case of the number of years spent in schooling, in the UK the age at which full-time education was finished is used. Thus, the category of less than an undergraduate degree refers to those who left full time education at age 20 or earlier, while the category of with at least an undergraduate degree refers to those who left full time education at age 21 or later. In France and the US, the two categories are defined using the highest qualification declared.

B, table (1.3)) and the estimate of the cohort size when the dependent variable is the number of years spent in schooling (the estimate reported in the first row of Panel A, table (1.6)) to the already found returns to education in the literature. Specifically, the literature identifies an increase in annual adult men income of 7-12% from spending one extra year in school (Bound et al. (1995), Angrist and Krueger (1992), Ashenfelter and Krueger (1994)). So, a ratio of the coefficients, as identified in the current paper, larger than 7-12% stands as a proof that the impact of cohort size on wages is not solely explained by worse educational attainment. In the US, for men, the ratio of the coefficients is 16.55%[18], while for French men it is around 3.4.%[19]. Therefore, it seems that it is only in the US that the impact of cohort size persists for wages, while in France the impact identified before is only due to the lower educational levels attained by the baby boomers.

Although one might question the fact that in the UK there is no effect of the cohort size on wages, while in the US wages are found to be negatively affected, it is worth mentioning that the stronger union power in the UK and the collective bargaining wage setting framework in the US, support these results.

1.6.2 Immigration

Our baseline analysis shows that larger cohort sizes increase unemployment rates in all three countries and decrease wages in the US, while also decreasing educational attainment across all three countries. In this section the analysis goes one step forward, in considering how immigration rates respond to the size of the total number of live births registered by that specific birth cohort. If large cohorts struggle on the job market, it is expected that there would be lower incentives for immigrants to move in any of the three countries considered, given that one of the main underlining reasons for migration is to join the labour market.

In order to test this hypothesis, equation (1.11) is estimated when the dependent variable is the immigration rate registered by cohort c at survey time t. We define immigration rates as the ratio between the number of immigrants born in year c and the total number of natives and immigrants born in year c and who were surveyed and present at time t in each country. One potential issue with this definition is if immigration rate is a function of the current cohort size, which is defined in our main analysis as the number of natives born in year c and surveyed at time t. Thus, we report only the reduced form estimates.

[18] Identified as the ratio between -0.397 and -2.398

[19] Identified as the ratio between -0.175 and -4.469

That is, the independent variable is the total number of live births registered in year c in each country.

Panel A of table (1.7) shows that, on average, larger number of births in a given cohort decrease immigration rates across countries and gender, when all educational levels are considered. Panel B restricts the sample only to individuals who have at least a high school or some college. In particular, among men, a 10 percentage points increase in the number of live births triggers a statistically significant decrease in the immigration rates of around 0.015 percentage points, 0.019 percentage points and 0.016 percentage points in France, the UK and US respectively. As for women, the magnitudes are slightly lower in France and the US, but larger in the UK, compared to men. This finding supports the results found in table (1.6), that most of the baby boomers do not have an undergraduate degree and also sheds light on the fact that the higher competition on the labour market at lower educational levels discourages immigrants of similar educational levels.

1.7 Conclusion

In this paper we study the effect of the cohort size on labour market outcomes. According to a simple supply-and-demand framework, demographic bulges increase the labour supply and cause an increase in the unemployment rates and/or a decrease in the real wages. Using variation in the population age structure registered in France, the UK and the US we test this hypothesis empirically.

The results obtained in an instrumental variable estimation show that the cohort size measured at age 25 has a statistically significant positive impact on unemployment rates. The impact has the highest magnitude among women in all three countries under analysis, with the largest increase in France. When the cohort size is measured at different ages, the magnitude of the effect decreases, becoming statistically insignificant when measured at age 40, at ones' late 40s and at age 50, in the UK, the US and France, respectively. We interpret this heterogeneity in results within and across countries due to different female labour market attachment at different ages and across countries.

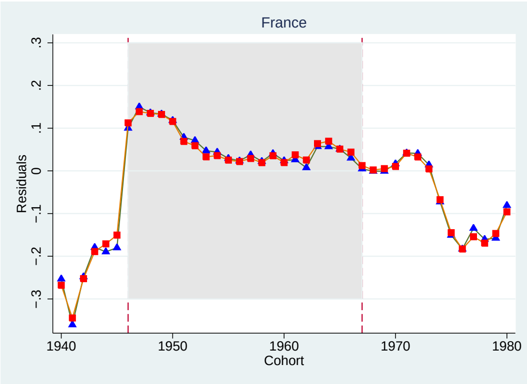
When the impact of the cohort size on wages is investigated, we find that even though initially it seems that it is mainly in the US and France that wages are negatively affected, once it is accounted for the fact that larger generations have lower levels of education, it is only in the US that wages are negatively affected by the cohort size measured at age 25. The impact on wages is larger in magnitude for men and lasts longer for men compared to women.

The paper also considers two potential mechanisms through which individuals can mitigate the identified negative effects of the cohort size on labour market outcomes: education and migration. Firstly, we find that individuals belonging to large generations spend less time in education, being more likely to finish education with less than an undergraduate degree, maybe due to the lack of response in the supply of university places. Moreover, we also show that immigration rates are negatively affected by the cohort size, measured as the number of live births. This underlines that the higher competition on the labour market at lower educational levels discourages immigrants of similar educational levels. Thus, we can conclude that baby boom generations face tougher labour market outcomes and are not able to mitigate their entry into the labour market through extending their time in education. These bleak outcomes discourage flows of similarly educated immigrants.

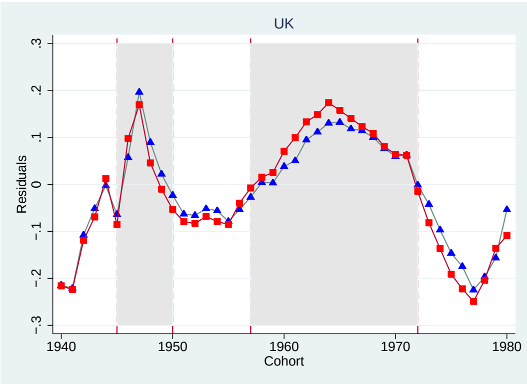
Tables and figures

FIGURE 1.1: Baby boom cohorts

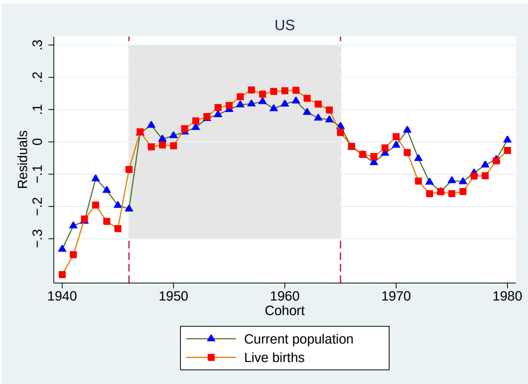
Panel A: France



Panel B: UK



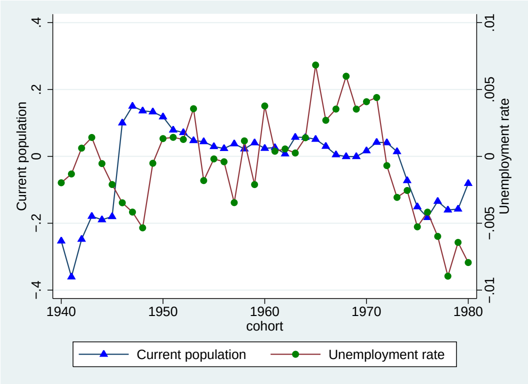
Panel C: US



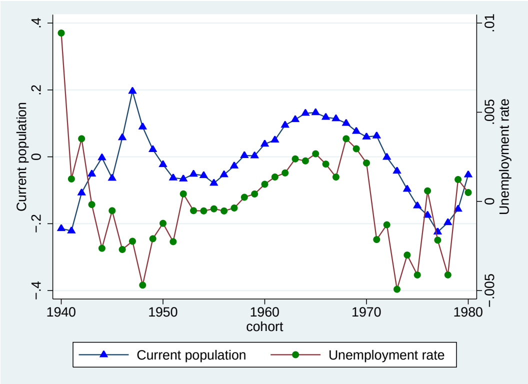
Notes: The line corresponding to the current population (includes only natives) plots the residuals from running equation (1.9), with standard errors clustered at cohort level. The line corresponding to the live births, plots the residuals of equation (1.10), with standard errors clustered at cohort level. The shaded areas identify the baby boom generations. Source: Author's calculations using the French Labour Force Survey, 1990-2012; the British Labour Force Survey, 1993-2013; the US March Current Population Survey, 1994-2013

FIGURE 1.2: Correlation between cohort size and unemployment rates

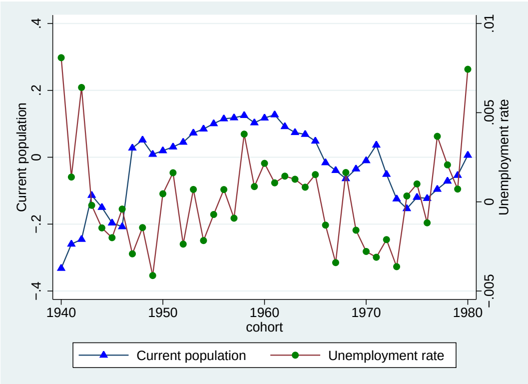
Panel A: France



Panel B: UK



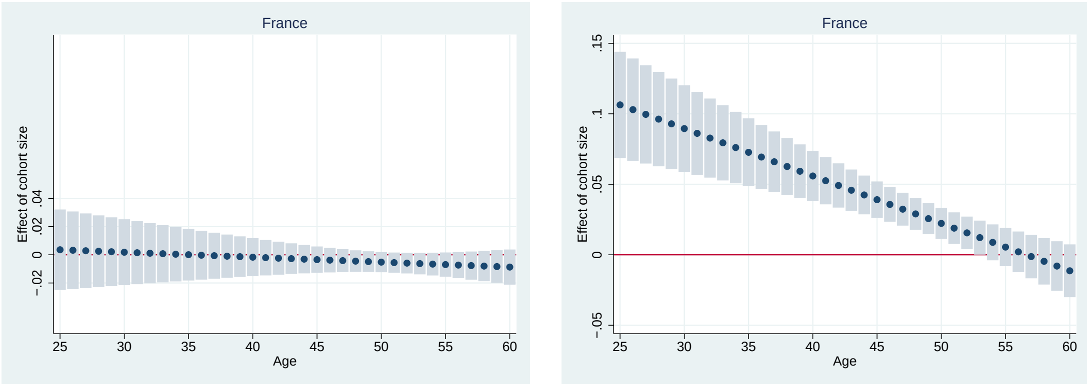
Panel C: US



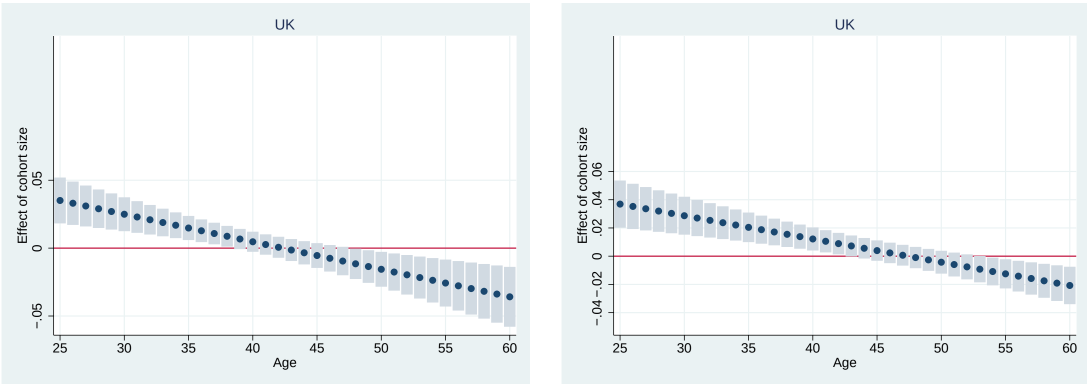
Notes: The figures plot the residuals of the regression of either the unemployment rate or the population against age and time fixed effects, clustering the standard errors at cohort level. Only natives are included in the calculations of the current population. Source: Author's calculations using the French Labour Force Survey, 1990-2012; the British Labour Force Survey, 1993-2013; the US March Current Population Survey, 1994-2013

FIGURE 1.3: Persistency of impact on unemployment rates

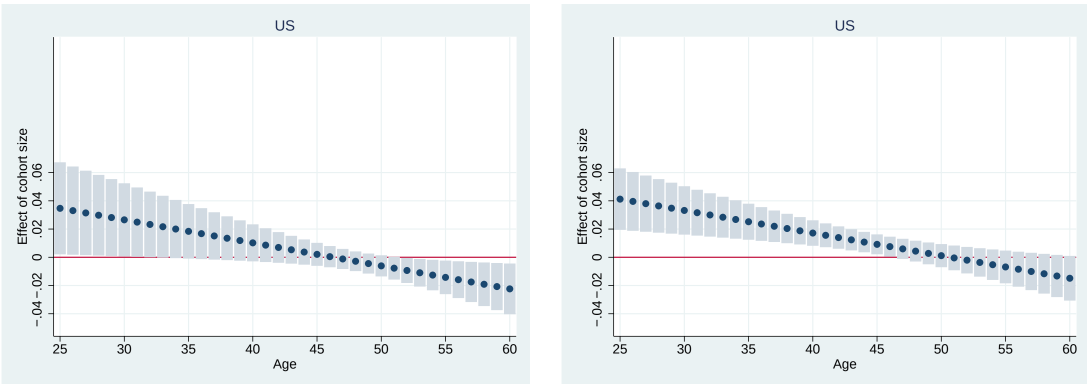
Panel A: France



Panel B: UK



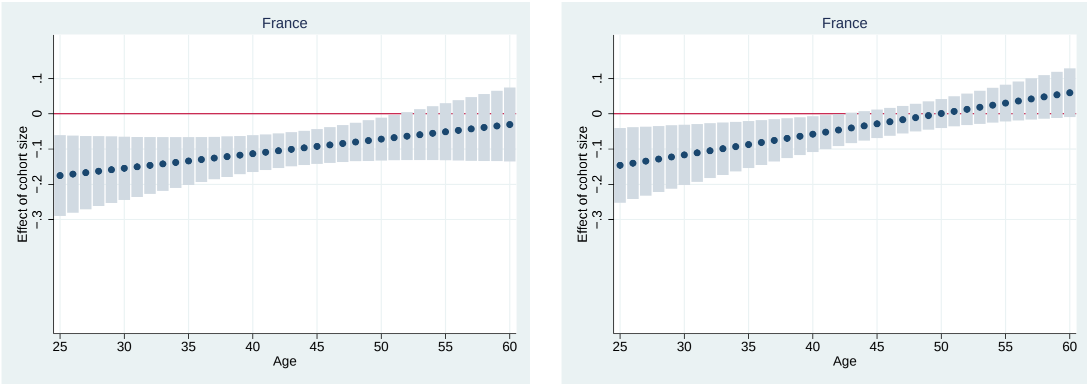
Panel C: US



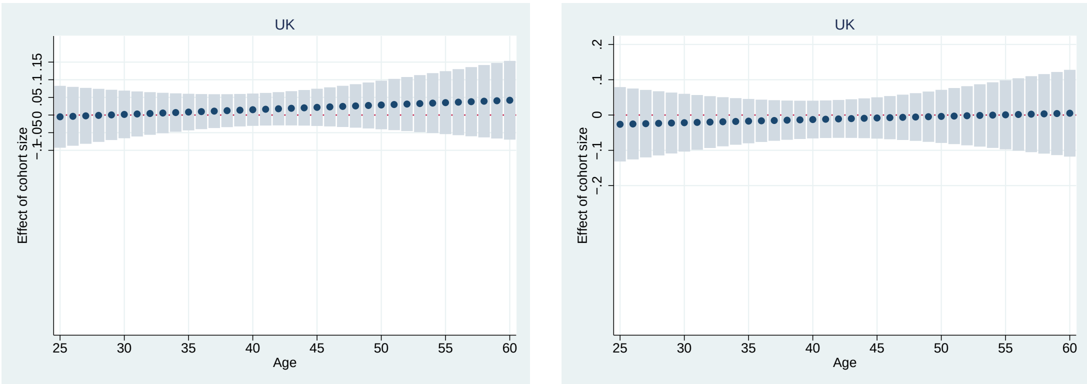
Notes: Each point represented in the graph is the estimate of $\beta_1 + \beta_2(a-25)$ obtained in the IV estimation of equation (1.11) where a is each age between 25 and 60. It identifies the impact of the cohort size at that specific age a on unemployment rates. Standard errors are clustered at cohort level. The shaded areas identify the limits of the 95% confidence intervals. Source: Author's calculations using the French Labour Force Survey, 1990-2012; the British Labour Force Survey, 1993-2013; the US March Current Population Survey, 1994-2013

FIGURE 1.4: Persistency of impact on weekly wages

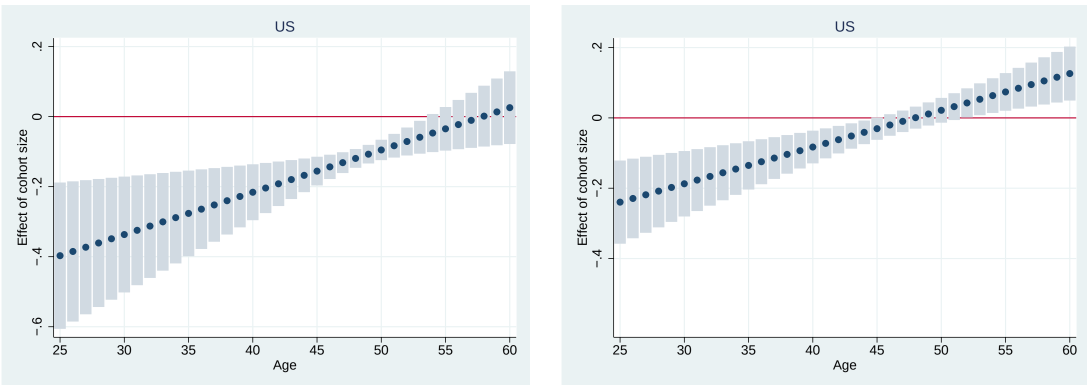
Panel A: France



Panel B: UK



Panel C: US



Notes:Each point represented in the graph is the estimate of $\beta_1 + \beta_2(a-25)$ obtained in the IV estimation of equation (1.11) where a is each age between 25 and 60. It identifies the impact of the cohort size at that specific age a on the logarithm of weekly wages. Standard errors are clustered at cohort level. The shaded areas identify the limits of the 95% confidence intervals. Source: Author's calculations using the French Labour Force Survey, 1990-2012; the British Labour Force Survey, 1993-2013; the US March Current Population Survey, 1994-2013

TABLE 1.1: Descriptive statistics

	France		UK		US	
	Mean	SD	Mean	SD	Mean	SD
Panel A. All						
Unemployment rate	0.075	0.027	0.047	0.018	0.049	0.019
ln Weekly wages	5.857	0.131	6.035	0.138	6.709	0.278
Number of years in schooling	9.744	1.062	12.227	0.627	13.774	0.256
Panel B. Men						
Unemployment rate	0.063	0.025	0.053	0.023	0.054	0.023
ln Weekly wages	5.928	0.156	6.126	0.146	6.815	0.271
Number of years in schooling	9.806	0.839	12.269	0.581	13.760	0.221
Panel C. Women						
Unemployment rate	0.089	0.033	0.040	0.016	0.044	0.017
ln Weekly wages	5.743	0.103	5.840	0.160	6.560	0.304
Number of years in schooling	9.685	1.299	12.187	0.678	13.783	0.331

Notes: The number of observations by country are: France - 745, the UK - 692, the US - 663. Wages are expressed in 1998 US dollars. Averages and standard deviations are reported.

TABLE 1.2: First Stage

	France	UK	US
	(1)	(2)	(3)
Panel A: ln Cohort size			
ln Live births	1.074***	0.970***	0.970***
	(0.037)	(0.055)	(0.114)
Panel B: ln Cohort size *(age-25)			
ln Live births*(age-25)	1.134***	0.925***	1.002***
	(0.043)	(0.135)	(0.097)
F statistic	1150.256	53.110	118.914
Observations	745	692	663
Time FE	X	X	X
Age FE	X	X	X

Notes: Panel A reports the estimate of $ln\ Z_{act}$ when the outcome variable is the logarithm of the cohort size. Panel B reports the estimate of $ln\ Z_{act}\ (a-25)$ when the outcome variable is the interaction between logarithm of the cohort size and the difference between age and 25. Robust standard errors clustered at cohort level in parentheses. F statistic is based on the Kleinbergen-Paap Wald F statistic. *** p<0.01, ** p<0.05, * p<0.1

TABLE 1.3: Effect on labour market outcomes

	Men						Women					
	France		UK		US		France		UK		US	
	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Unemployment rate												
ln Cohort size	0.005	0.004	0.027***	0.035***	0.019	0.035**	0.091***	0.106***	0.035***	0.037***	0.036***	0.041***
	(0.013)	(0.013)	(0.008)	(0.008)	(0.014)	(0.015)	(0.017)	(0.018)	(0.008)	(0.008)	(0.010)	(0.010)
ln Cohort size*(age-25)	-0.000	-0.000	-0.002***	-0.002***	-0.001*	-0.002**	-0.003***	-0.003***	-0.002***	-0.002***	-0.001***	-0.002***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
F statistic		1150.256		53.110		118.914		1150.256		53.110		118.914
Observations	745	745	692	692	663	663	745	745	692	692	663	663
Panel B: ln Weekly wage												
ln Cohort size	-0.161***	-0.175***	0.000	-0.005	-0.362***	-0.397***	-0.158***	-0.146***	-0.017	-0.026	-0.220***	-0.240***
	(0.054)	(0.054)	(0.037)	(0.041)	(0.085)	(0.098)	(0.049)	(0.050)	(0.047)	(0.049)	(0.048)	(0.055)
ln Cohort size*(age-25)	0.003	0.004	0.002	0.001	0.011***	0.012***	0.006***	0.006***	0.000	0.001	0.010***	0.010***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
F statistic		1150.256		53.110		118.914		1150.256		53.110		118.914
Observations	745	745	692	692	663	663	745	745	692	692	663	663
Time FE	X	X	X	X	X	X	X	X	X	X	X	X
Age FE	X	X	X	X	X	X	X	X	X	X	X	X

Notes: Regressions (1) - (6) refer to men, while regressions (7) - (12) refer to women. The outcome variable in panel A is the unemployment rate. The outcome variable in panel B is the logarithm of real weekly wages. The reported independent variables are the logarithm of the cohort size, the logarithm of the cohort size interacted with (age-25) Robust standard errors clustered at cohort level in parentheses. F statistic is based on the Kleinbergen-Paap Wald F statistic. *** p<0.01, ** p<0.05, * p<0.1

TABLE 1.4: Effect on labour market outcomes - weighted regressions

	Men						Women					
	France		UK		US		France		UK		US	
	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Unemployment rate												
ln Cohort size	0.006	0.005	0.031***	0.039***	0.016	0.031**	0.091***	0.108***	0.037***	0.039***	0.033***	0.040***
	(0.014)	(0.014)	(0.008)	(0.008)	(0.014)	(0.016)	(0.018)	(0.018)	(0.008)	(0.008)	(0.011)	(0.010)
ln Cohort size*(age-25)	-0.000	-0.000	-0.002***	-0.002***	-0.001	-0.001**	-0.003***	-0.003***	-0.002***	-0.002***	-0.001***	-0.002***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
F statistic		833.097		67.695		113.540		833.097		67.695		113.540
Observations	745	745	692	692	663	663	745	745	692	692	663	663
Panel B: ln Weekly wage												
ln Cohort size	-0.169***	-0.188***	-0.002	-0.008	-0.336***	-0.362***	-0.146**	-0.131**	-0.022	-0.027	-0.206***	-0.229***
	(0.057)	(0.055)	(0.036)	(0.039)	(0.089)	(0.103)	(0.056)	(0.054)	(0.049)	(0.052)	(0.051)	(0.059)
ln Cohort size*(age-25)	0.004	0.005*	0.001	0.001	0.010**	0.011**	0.006**	0.005**	0.001	0.001	0.009***	0.010***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
F statistic		833.097		67.695		113.540		833.097		67.695		113.540
Observations	745	745	692	692	663	663	745	745	692	692	663	663
Time FE	X	X	X	X	X	X	X	X	X	X	X	X
Age FE	X	X	X	X	X	X	X	X	X	X	X	X

Notes: Regressions (1) - (6) refer to men, while regressions (7) - (12) refer to women. The outcome variable in panel A is the unemployment rate. The outcome variable in panel B is the logarithm of real weekly wages. The reported independent variables are the logarithm of the cohort size and the logarithm of the cohort size interacted with (age-25). The weights are the number of men/women in each age-cohort-year cell. Robust standard errors clustered at cohort level in parentheses. F statistic is based on the Kleinbergen-Paap Wald F statistic. *** p<0.01, ** p<0.05, * p<0.1

TABLE 1.5: Effect on labour market outcomes - state of the economy at age 25

	Men						Women					
	France		UK		US		France		UK		US	
	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)	(OLS)	(IV)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Unemployment rate												
ln Cohort size	0.005	0.003	0.027***	0.035***	0.019	0.035**	0.089***	0.106***	0.035***	0.036***	0.035***	0.041***
	(0.013)	(0.013)	(0.008)	(0.008)	(0.014)	(0.015)	(0.019)	(0.019)	(0.008)	(0.008)	(0.009)	(0.010)
ln Cohort size*(age-25)	-0.000	-0.000	-0.002***	-0.002***	-0.001*	-0.002**	-0.003***	-0.003***	-0.002***	-0.002***	-0.001***	-0.002***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
GDP	-0.035	-0.036	-0.002	0.002	0.012	0.011	-0.121**	-0.121***	-0.008	-0.008	0.030	0.029
	(0.025)	(0.024)	(0.018)	(0.017)	(0.019)	(0.018)	(0.051)	(0.046)	(0.017)	(0.016)	(0.021)	(0.020)
F statistic		1281.842		52.30		122.037		1281.842		52.30		122.037
Observations	745	745	692	692	663	663	745	745	692	692	663	663
Panel B: ln Weekly wage												
ln Cohort size	-0.150**	-0.172***	-0.008	-0.014	-0.365***	-0.399***	-0.153***	-0.144***	-0.015	-0.015	-0.219***	-0.239***
	(0.056)	(0.060)	(0.037)	(0.042)	(0.086)	(0.099)	(0.050)	(0.052)	(0.047)	(0.047)	(0.048)	(0.055)
ln Cohort size*(age-25)	0.004	0.005*	0.002	0.002	0.011***	0.012***	0.006***	0.006***	0.000	0.000	0.010***	0.010***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
GDP	0.615**	0.630***	-0.173**	-0.170**	0.152	0.154	0.295*	0.305**	0.051	0.051	-0.044	-0.044
	(0.229)	(0.214)	(0.081)	(0.077)	(0.111)	(0.104)	(0.163)	(0.154)	(0.144)	(0.144)	(0.128)	(0.119)
F statistic		1281.842		52.30		122.037		1281.842		52.30		122.037
Observations	745	745	692	692	663	663	745	745	692	692	663	663
Time FE	X	X	X	X	X	X	X	X	X	X	X	X
Age FE	X	X	X	X	X	X	X	X	X	X	X	X

Notes: Regressions (1) - (6) refer to men, while regressions (7) - (12) refer to women. The outcome variable in panel A is the unemployment rate. The outcome variable in panel B is the logarithm of real weekly wages. The reported independent variables are the logarithm of the cohort size, the logarithm of the cohort size interacted with (age-25) and the growth rate of real GDP. Robust standard errors clustered at cohort level in parentheses. F statistic is based on the Kleinbergen-Paap Wald F statistic. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 1.6: Mechanism: education

	Men			Women		
	France	UK	US	France	UK	US
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Average number of years in school						
ln Cohort size	-4.469***	-1.761***	-2.398***	-2.739***	-1.688***	-1.761***
	(0.903)	(0.193)	(0.413)	(0.734)	(0.209)	(0.307)
ln Cohort size*(age-25)	0.152***	0.056***	0.106***	0.095***	0.048***	0.087***
	(0.035)	(0.011)	(0.018)	(0.028)	(0.009)	(0.012)
F statistic	1150.256	53.110	118.914	1150.256	53.110	118.914
Observations	745	692	663	745	692	663
Panel B: Proportion- less than an undergraduate degree						
ln Cohort size	0.287***	0.381***	0.372***	0.284***	0.462***	0.339***
	(0.065)	(0.033)	(0.076)	(0.086)	(0.028)	(0.059)
ln Cohort size*(age-25)	-0.008***	-0.012***	-0.014***	-0.007**	-0.013***	-0.013***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)
F statistic	1150.256	53.110	118.914	1150.256	53.110	118.914
Observations	745	692	663	745	692	663
Time FE	X	X	X	X	X	X
Age FE	X	X	X	X	X	X

Notes: Regressions (1) - (3) refer to men, while regressions (4) - (6) refer to women. In Panel A the outcome variable is the average number of years in school. In Panel B the outcome variable is the proportion of people who have less than an undergraduate degree. The reported independent variables are the logarithm of the cohort size and the logarithm of the cohort size interacted with (age-25). Robust standard errors clustered at cohort level in parentheses. F statistic is based on the Kleinbergen-Paap Wald F statistic. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 1.7: Mechanism: immigration

	Men			Women		
	France	UK	US	France	UK	US
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All schooling levels						
ln Live births	-0.127***	-0.235***	-0.118***	-0.097***	-0.241***	-0.110***
	(0.024)	(0.018)	(0.028)	(0.024)	(0.019)	(0.034)
ln Live births*(age-25)	0.002**	0.005***	0.001	0.002**	0.005***	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	745	692	663	745	692	663
Panel B: Less than an undergraduate degree						
ln Live births	-0.145***	-0.186***	-0.163***	-0.134***	-0.200***	-0.151***
	(0.023)	(0.014)	(0.029)	(0.026)	(0.015)	(0.034)
ln Live births *(age-25)	0.002***	0.003***	0.002*	0.004***	0.004***	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	745	692	663	745	692	663
Time FE	X	X	X	X	X	X
Age FE	X	X	X	X	X	X

Notes: Regressions (1) - (3) refer to men, while regressions (4) - (6) refer to women. Panel A estimates the impact on immigration rates for all individuals. Panel B estimates the impact on immigration rates only for individuals who have less than an undergraduate degree. The reported independent variables are the logarithm of the number of live births in each cohort and the logarithm of the number of live births in each cohort interacted with (age-25). Robust standard errors clustered at cohort level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Chapter 2

The impact of foreign students enrolled in British universities

2.1 Introduction

In the current globalised economy driven by human capital, the number of individuals pursuing higher education abroad continues to surge. Even though the competition for the smartest minds from all over the world is stronger with the recent openness of more universities to foreign students, countries like the UK, which have universities with abiding reputation, are still a top destination for students studying abroad.[1] Aggregated figures show that over the last years the number of non-UK domiciled students enrolled as first year full-time undergraduate students in all British universities hiked by 75% from 37,515 to 65,805 between academic years 2000/01 and 2009/10.[2]

Economic theory suggests that in a world with fixed university places increases in the number of mobile undergraduate students decreases the number of competing undergraduate native students enrolled in universities. However, universities have some flexibility in altering the supply of places, and, at least in the UK, they have been expanding over time.[3] So, rather than seeing a mechanical relationship in which one extra foreign student displaces one native, fluctuations in university places make it difficult to predict what is the effect of increases in demand from non-UK domiciled students. On the one hand, universities have limited resources and extra foreigners could crowd out natives;

[1]According to UNESCO data the UK was the second top destinations for students studying abroad in 2013 enrolling 10% and following the US which enrolled 19%.

[2]In this paper the group of the EU and non-EU domiciled students is referred to interchangeably as non-UK domiciled students, foreign students or mobile students.

[3]According to the Higher Education Statistical Agency, between 2000/01-2009/10 the total number of first year full time undergraduate students enrolled in all British universities increased from 383,365 to 516,480.

on the other hand, these extra foreigners could become an additional source of income for universities through the tuition fees they pay.[4] In this case mobile students could crowd in native students as universities may invest these extra financial resources to expand, by creating more teaching facilities or hiring extra teaching staff. Moreover, even if tuition fees are the same for everyone, independently of their domicile, universities may want to increase enrolment rates of mobile students, as they bring cultural diversity on campus which could contribute to the enhancement of one's university experience. One additional aspect is the quality of the students enrolled: if foreign applicants tend to be of higher ability than natives on average, universities may have incentives to enrol more non-UK domiciled students in order to increase their own competitiveness.

Thus, as the direction of the impact is ambiguous, in this paper we empirically assess how the large inflows of foreign undergraduate students attending British universities has affected the enrolment of native students. Specifically, we combine very rich individual level administrative data on eight cohorts of English students and on non-UK domiciled students to analyse the overall effect. Then, we extend the analysis by investigating how this increased competition from foreigners has altered the distribution of native students enrolled in universities. We focus on analysing if there are unequal effects between natives by their academic performance in pre-university national level exams to investigate if it is the more or the less able students who experience a greater effect. Moreover, we also distinguish between natives from different demographic groups in order to understand if it is the poor or the richer native students who benefit from or are negatively impacted by the larger influx of foreign students.

We use an instrumental variable (IV) approach to account for the fact that foreign and native students are subject to similar university specific demand shocks, such as the fact that a university may become more attractive to both categories of students because it is expanding. The used instrument parallels that proposed by Card (2001), which is widely used in the labour economics literature. In particular, we use historical shares of students from a sending country enrolled into a university combined with current national changes in the stock of students from this country to instrument the current flows of foreign undergraduate students attending a university. This supply-push component of recent foreign inflows to a particular university, which is arguably exogenous to university demand conditions, allows us to identify the causal effect of non-UK domiciled inflows of students in the presence of unobserved university demand shocks on various university related outcomes for natives. Moreover, as since the introduction of tuition

[4]In the UK students domiciled outside the EU pay higher tuition fees compared to native and EU students. The British government estimated that non-EU students contributed £3.9bn in tuition fees after scholarships and £6.3bn in living expenses in 2011/12 (BIS (2013)).

fees in 1998/99 the funding system of British universities has undergone a series of changes such as moving from upfront to deferred tuition fees, the introduction of tuition fee loans or changes in maintenance grants and loans, we employ an estimation with flexible controls for time fixed effects to ensure our empirical analysis is not affected by these changes.

Our IV results show that even though there is no statistically significant impact overall, there is variation in the effect by native groups. We find the top performing native students are crowded in by foreign students: a 1% increase in the number of foreign students triggers an increase in the number of native students with grades above the median GCSE English grades and above the median GCSE Mathematics grades by 0.14% and 0.23% respectively. Additionally, a 1% increase in enrolled undergraduate foreigners increases the number of male native students by 0.13%. Our findings show a crowding in of natives from the top distribution of the income, although the effect is differential only at the 10% significance level. As for ethnic origins, we show that UK minorities whose first language is not English and mainly those of Asian origins are also more likely to enrol into university. Our distributional analysis by natives' demographic composition complements our distributional analysis by natives' ability, as there is a large literature that shows that in England students of Asian origins tend to score better than UK-born white pupils in the exam taken at the end of compulsory school (Dustmann et al. (2010), Rutter (2016), Strand (2014), Hutchinson et al. (2016)).

With the introduction of tuition fees in 1998/99 British universities have slowly transitioned to a free market. In particular, although tertiary education institutions have quotas for the number of EU domiciled and native students they enrol, there are no restrictions on the number of non-EU domiciled students they enrol.[5] As these non-EU students pay considerably higher tuition fees compared to natives and EU students, universities have more incentives to attract them, potentially using these extra resources to attract more top performing native students. Thus, we investigate this potential mechanism, but we find no evidence that the effect is triggered by the higher resources available to universities through the tuition fee paid by non-EU domiciled students. We further investigate whether foreign students crowd in top native students as their enrolment increases the quality of the university attended and thus increasing its appeal. We find limited evidence that foreigners increased slightly the ranking of universities.

[5]The transition to a free market is more striking for the period after the increase in tuition fees in 2012. Tertiary education institutions receive considerable less direct funding from the government, have no restrictions on student numbers anymore and also charge considerably higher tuition fees. Thus, universities are competing even more for additional funds and have more incentives to attract higher payers of tuition fees, who tend to be foreign students.

The rest of the paper is structured as follows. Section 2.2 presents the literature review. Section 2.3 describes the British education system. Section 2.4 details the data. Section 2.5 explains the estimation strategy and offers solutions to potential estimation challenges. Section 2.6 reports the results, while section 2.7 tests the robustness of the results. Section 2.8 explores the mechanisms. Section 2.9 discusses the results and concludes.

2.2 Literature review

The effect of immigrant inflows on receiving markets and natives has generated a large debate in the literature. The impact of immigration on labor market outcomes in particular has proven a controversial issue. On the one hand, Card (1990), Card (2001), Card (2005), Manacorda et al. (2012) and Ottaviano and Peri (2012) argue that immigration has had little and often insignificant effects on native workers' wages and employment rates (Dustmann et al. (2005)); on the other hand, Borjas et al. (1996) and Borjas (2003) find a pronounced negative effect on natives' wages.

Surprisingly, however, the impact of immigration on the higher education system has not been largely studied. Increasing enrolment rates of foreign students can alter the educational opportunities of natives: one extra foreigner could displace natives from tertiary education or even discourage natives from pursuing degrees popular among foreign students, especially if after graduation mobile students are very likely to join the labour market in the host country. Thus, the issue of crowding out effects has significant policy implications in the current context of increasing numbers of students pursuing a degree abroad.

To our knowledge, the only other study analysing the issue of foreign students in the UK is the work by Machin and Murphy (2014). They use aggregated data on enrolment in British universities to examine whether non-EU domiciled students crowd out native students. Their findings suggest that there is no overall effect among undergraduate native students, but that taught-postgraduate students are crowded in. The authors find that the higher tuition fees paid by these non-EU students help universities to attract more native students. In our paper we revisit this overall effect for undergraduate students and we further analyse the distributional effects of the impact of foreign students on natives by using detailed individual level data. We mainly focus on the academic performance and the demographic structure of enrolled natives. Moreover, through analysing how the ethnic composition of the native student body is affected by the inflow of foreign students, we contribute to the understanding of how the integration of

British-born minorities, different in culture and religion, has responded to increasing numbers of mobile students in the UK. In this sense our paper is related to the study of Dustmann et al. (2003) who argue that labour market outcomes of ethnic minority individuals who are born in the UK are better than those of immigrants (relative to the UK-born Whites), but that many communities are still disadvantaged compared to the White UK-born population.

The other studies at tertiary education level focus mainly on the US and find modest evidence that increasing competition from foreign students has affected the university opportunities of natives. Jackson (2015) uses data from the US Census between 1970 and 2000 to find that state-level increases in the number of immigrant university students do not reduce enrolment rates of US natives. Furthermore, Borjas (2007) analyses enrolment trends in US graduate programs between 1978 and 1998 and finds no crowding out effect on average, although there is heterogeneity in the impact across ethnic groups with White native men being negatively affected by the large number of foreign students. Hoxby (1998) studies whether immigrants push disadvantaged American natives out of higher education, by exploiting a policy change in the fee structure within the Californian higher education system between 1986 and 1992. The results show that Black and Hispanic students are displaced by less disadvantaged foreign-born pupils.

In another related paper, Kato and Sparber (2013) analyse the effect of a restriction in visas available to foreign-born workers on the quality of undergraduate applications that US universities receive from international students. Whilst on a different research question, the study shows that the decrease in the number of issued visas triggered a drop in the number of applications of high-ability foreign students. In our paper we will also aim to investigate how sensitive the average quality of native undergraduate students enrolled in British universities is to larger inflows of foreign students.

Studies at other levels of education have found both overall and distributional statistically significant effects of immigrants on natives. Gould et al. (2009) use the mass migration inflow of immigrants in Israel in the 1990s to examine the impact of immigrant concentration in elementary school on the long-term academic outcomes of native students in high school. The results point to lower likelihood of natives passing the high school matriculation examination that are key for university enrolment. Hunt (2012) examines the impact of immigration on natives' high school completion in the United States. The author finds that native-born Black pupils, especially, are more encouraged to complete high school in order to avoid competing with immigrant high school dropouts in the labour market. Betts (1998) studies whether immigration affects the probability of high school graduation of American-born minorities. Results suggest that the native-born Blacks are more likely to have lower retention rates. Our paper is the first to investigate the differential effects by natives' ethnic characteristics in the UK at the tertiary education level.

2.3 Institutional setting

In England, full-time education is compulsory for all children aged between 5 and 16 years old and it is organised in five Key Stages (KS). A national level examination, called General Certificate of Secondary Education (GCSE), marks the end of compulsory education. Students have the freedom to choose which and how many subjects to take, but everyone takes written exams in around ten different subjects and sits the GCSE English and Mathematics.

At the end of compulsory education students decide to either finish formal education or continue their studies for two more years, choosing between a vocational or an academic track. At the end of these two years, most English students who want to pursue a bachelor degree and who are by now 18/19 years old take a national level exam, called the General Certificate of Education Advanced Level (A-levels), in three or four subjects. The choice of subjects tends to be closely related to one's university degree preferences and university admissions are mainly determined by the scores obtained at the A-levels.[6]

When applying to a British university students choose specific fields of study and their degree can vary in length based on the location and the subjects studied, with most lasting three years in England, Wales and Northern Ireland and four years in Scotland.[7]

Although there is a large number of universities/subjects to choose from, institutions compete to attract students. Every summer a number of British university league tables are published.[8] Through providing information on the quality of each university or/and various field specific departments based on a set of objective criteria they aim to help

[6]Some universities like Cambridge or Oxford also ask prospective students to attend an interview as part of the admission process.

[7]The application process is centralised and each students applies through UCAS to up to five university-field of study groups. Applications are analysed separately by each institution-department and offers are made conditional on the grades obtained at the A-level exam, which is taken after the university admission process is ended. Students need to choose their top two preferences of the offers received before sitting the A-level and if they meet the grade requirements they can enrol into university. Students that did not meet the thresholds imposed by either of their two options may still find a free spot at university which did not fill in all their positions by going into clearing.

[8]The Times university rankings were first published in 1992, the Sunday Times introduced theirs in 1998, the Guardian followed in 1999 and the Complete University Guide (the Independent) in 2007.

prospective students in choosing the universities and subjects to apply for. The importance of British university league tables to prospective students has been documented in the literature, with Soo (2013); Broecke (2015) and Gibbons et al. (2015) finding that improvements in university rankings are associated with increases in the number of applications and underlining that students sort into universities based on university ranking.

At admission, British universities distinguish between students based on domicile, splitting them in two main categories: home and overseas students. The former group includes all students domiciled in the UK or in a EU country, while the latter refers to all students domiciled in countries which are not part of the EU. This distinction is crucial as the two groups are subject to different regulation in terms of tuition fees levels, available places and funding opportunities. In a nutshell, universities have upper boundaries for tuition fees levels as well as student number caps imposed by the government for home students, but no regulation is in place for overseas students.[9]

In 1998, universities in the UK started charging their undergraduate students upfront mean-tested annual tuition fees of up to £1,000.[10] In 2006 universities in England, Wales and Northern Ireland introduced variable fees, with each institution having the discretion over the amount of fees they charged up to a maximum of £3,000 for home students. In the following years the maximum level was inflation-indexed. This fee regulation applied to all home students who were also eligible to apply for tuition fees loans offered by the Student Loan Company and payable after graduation in instalments, once earnings have reached £15,000 annually.[11] A different tuition fee regime has been in place in Scotland since 2001 due the devolution, but in general, universities tended to charge much larger tuition fees for non-EU students. For instance in 2011, according to The National Survey of UK Tuition, undergraduates in universities in England, Wales and Northern Ireland paid on average an annual fee of £3,375.[12] Yet, undergraduates from countries outside the EU were charged fees ranging from £6,000 to £23,000 depending on the university and/or the type of degree pursued.[13]

[9]This refers to the period under analysis in our paper: academic years 2004/05-2011/12.

[10]Students were exempt from fees if their families earned less than £23,000 per year and were charged reduced fees on a decreasing scale if their families earned between £23,001 and £35,000 per year. Students whose families earned at least £35,001 were charged full tuition fees.

[11]Source: Student Loan Repayment. Website: www.studentloanrepayment.co.uk

[12]Scottish universities imposed no fee on students from Scotland or the EU studying full time on their first degree, and £1,820 on English, Welsh and Northern Irish residents.

[13]Source: The complete university guide. Website: http://www.thecompleteuniversityguide.co.uk

The university places available to home students (i.e. natives and EU students) were also regulated by the government during the period under analysis. Specifically, if universities went over or below the threshold of 3-5% of the student numbers proposed by government bodies, they will face funding penalties in subsequent years for their home students. Yet, the decision of how many non-EU students to enrol in a given year is mainly based on demand and the capacity of teaching as well as constraints imposed by the Home Office, due to visa restrictions, as each university has to become a sponsor and apply for a confirmation for acceptance for studies from the Home Office for each potential student not domiciled inside the European Economic Area.

Thus, in this diverse education system universities have the financial incentive to increase the number of overseas students as they pay larger tuition fees. This raise in available funds could help universities to invest in their teaching or research quality and increase their overall capacity and improving facilities. However, in the short run they face capacity constraints which could lead to displacement of the UK and/or EU students as well as penalties from the government.

2.4 Data

We use two main sources of individual level data. The linked National Pupil Dataset (NPD) - Higher Education Statistical Agency (HESA), which is jointly provided by the English Department for Education (DfE) and HESA, contains information on all British domiciled pupils who finished compulsory education in English state schools and pursued an undergraduate degree in a British university. The Student Record contains administrative information on all non-UK domiciled undergraduate students enrolled in a British university and it is provided directly by HESA.

Our main analysis focuses on eight cohorts of undergraduate students who enrolled in a British university between academic years 2004/05-2011/12. Additionally we also use data on foreign students between 1998/99-2003/04. In order to control for changes in the supply of places due to university merges, openings or closures, a balanced panel of universities which reported a positive number of enrolled students at undergraduate level over the period 1998/99-2011/12 is considered (See Appendix B for full details). This leads to 139 universities in total. In order to increase the precision of the estimation we group the 20 JACS fields of study identified in the data in five groups: Medicine, Dentistry and Allied Subjects; STEM; Social Sciences; Languages and History; Arts, Education, Other (See Appendix B for a detailed description).

We restrict the sample to first year full-time undergraduates. Table (2.1) summarises the data contained in the two data sets provided by HESA, which mainly refers to university related information, by domicile. Panel A shows the details for those UK domiciled students who finished their secondary education in an English state school between 2001/02-2008/09 and enrolled as first year undergraduate students in a British university between 2004/05-2011/12. When considering the quality of the university attended, only 19% are pursuing a degree in one of the 20 leading British universities which form the Russell Group.[14] Given that the cohorts of natives in our data are students who finished their compulsory secondary school in England, it is expected that majority enrol in an English university (96%). Regarding the field of study pursued, 29% enrol in Social Sciences, followed by those in Medicine, Dentistry and Allied subjects and STEM degrees, with shares of around 23% and 19%, respectively.

Panel B, shows that the share of students from non-EU countries represents around 61% of all non-UK domiciled students. Graph (2.1) shows that both types of students have registered increasing flows over time, while graph (2.2) exhibits that indeed the most representative are non-EU students, with the Chinese being by far the largest number on average. They are followed by those from Hong Kong, the US, France, Germany and Cyprus. Moreover, around 28% are enrolled in Russell group universities and 85% pursue a degree in an English university. As for the field of study, the largest share of foreigners pursue a degree in Social Sciences (approximately 42%) followed by around 23% enrolled in a STEM degree.

The NPD data provides additional information on natives: both demographic characteristics available in the annual census and the results at GCSE taken at the end of KS4.[15] On average, about half a million pupils finish secondary education in an English state school every year and they represent around 93% of all English pupils, the remaining being enrolled in independent schools. Out of these approximately 34% continue into the tertiary education level. Table (2.2) presents summary statistics of selected key variable for these students enrolled in higher education. Panel A shows information on student background characteristics at age 16. Out of all natives enrolled in university

[14]The Russell Group was formed in 1994 by 17 British research universities: University of Birmingham, University of Bristol, University of Cambridge, University of Edinburgh, Imperial College London, University of Leeds, University of Liverpool, London School of Economics and Political Science, University of Manchester, Newcastle University, University of Nottingham, University of Oxford, University of Sheffield, University of Southampton, University College London and University of Warwick. Cardiff University and King's College London became part of the group in 1998. Queen's University Belfast also joined the group in 2006. Since 2012 the group extended to include 24 universities, with the addition of Durham University, University of Exeter, Queen Mary University of London and University of York. Thus, in our paper we refer to the Russell Group as all 20 universities that formed the group before 2011/12.

[15]The School Census replaced the Pupil Level Annual Schools Census in 2006 for secondary schools.

approximately 55% are female. 80% of pupils are white and the largest minority group is represented by students of Asian origin (around 11.2%). 86% of UK-born students speak English as their first language. The data also identifies students eligible for free school meals, which is considered as a good proxy for family income: 7% of those enrolled in university were eligible for the free school meal at age 16.[16] The Income Deprivation Affecting Children Index (IDACI) is an index of poverty calculated by the Office of the Deputy Prime Minister, measuring the proportion of children under 16 years old that live in low income households within a local area. It is a continuous measure between 0 and 1, with higher values corresponding to students living in more impoverished areas. That is, children from worse off areas are less likely to enter university (17% on average).

Panel B shows details on the academic performance at the GCSEs. Students enrolled in university take on average 10 subjects at the GCSE level and have quite high grades, with 88% taking at least 5 A*-C. Moreover, they have high grades in English and Mathematics, the two compulsory subjects for all students, given that the mean grade for all their cohort sitting the GCSE is 0 (See appendix B for the conversion of the grades into numerical grades), and those that end up in tertiary education have on average a grade of around 0.7. Finally, students in university also attended better secondary schools, which have higher average test scores in English and Mathematics at the GCSE compared to the overall cohort for which the average standardised scores are 0.

In summary, this rich data allows us to follow natives through the education system and to analyse how the increasing inflow of non-UK domiciled students has affected the enrolment of these natives, as well as which categories of natives are more likely to be affected and how.

2.5 Empirical strategy

In this section we discuss the empirical strategy used to estimate the effect of the increasing inflows of foreign undergraduate students on English students. We begin by presenting the main estimation and then we describe the instrumental variable strategy used to control for the potential endogeneity of the flows of foreign students.

[16]This is a binary indicator of whether a pupil's family has claimed eligibility for free school meal. Only pupils from families that receive income benefits are eligible.

2.5.1 Main estimation

The goal of this paper is to estimate the effect of changes in the number of foreign students enrolled in British universities on UK domiciled students' enrolment rates. To do so, we estimate the following equation:

$$ln\ N_{ut} = \alpha_0 + \alpha_1 ln\ F_{ut} + d_t + d_u + d_u * t + \varepsilon_{ut} \tag{2.1}$$

where $ln\ N_{ut}$ is the natural logarithm of the total number of natives enrolled in university u in academic year t; $ln\ F_{ut}$ is the natural logarithm of the total number of non-UK domiciled enrolled; d_t and d_u are university and time fixed effects, respectively; $d_u * t$ captures the interaction between university fixed effects and a time trend; ε_{ut} is the residual.

The full array of fixed effects account for the university and time specific conditions that would bias results if omitted. The interaction between the university fixed effects and a time trend controls for time-varying university specific characteristics. Moreover, by controlling for this array of fixed effects we believe that our estimation is not affected by the various changes in the British higher education funding system which took place during the period covered in the analysis (for instance, increases in the level of tuition fees and changes in the nature of the fees from upfront to a deferred system or increases in maintenance grants and loans).[17]

We cluster standard errors at university level. The coefficient of interest is α_1. The inclusion of both the dependent and the independent variables in logarithmic forms allows the coefficient of interest, α_1, to be interpreted as an elasticity. An estimate of $\alpha_1 >= 0$ implies that extra foreign students crowd in natives across universities, while an estimate of $\alpha_1 < 0$ implies a crowding out effect.

[17] As briefly explained in section 2.3 and in much more detail in Chapter 3, section 3.3, various changes in the funding of higher education in the UK have been in place since the late 1990s. The first major change was the introduction of income contingent tuition fees in the academic year 1998/99, which forced to pay up to approximately £1,000 at the beginning of each academic year. The Higher Education Act 2004, effective from 2006/07, changed the regime again through the introduction of variable tuition fees. English, Welsh and Northern Irish universities had discretion over the level of the tuition fees charged, up to a maximum of £3,000 per annum (inflation indexed), with Scotland implementing different policies. Although these fees were not means tested, all native students were eligible to apply for tuition fee loans, independent of their economic situation and the value of the loan would cover the entire cost of tuition fee, payable in instalments, after graduation and once their income level exceeded £15,000 and the interest rate was very small, close to zero. During the period under analysis, the maintenance grants which were halved in 1998 and then abolished in 1999, were reintroduced in 2004/05. In addition, there have been a number of increases in means tested maintenance loans throughout the period of analysis.

When estimating equation (2.1) we assume that universities make centralized adjustments to student numbers across fields of study. However, when applying for an undergraduate degree in a British university, students choose the specific degree they want to pursue, not only the university. Thus, we further explore the variation in the numbers of students enrolled in each university - field of study group:

$$ln\ N_{uft} = \beta_0 + \beta_1 ln\ F_{uft} + d_u + d_f + d_t + d_{tf} + d_{uf} + d_{tu} + d_{uf} * t + \varepsilon_{uft} \tag{2.2}$$

where $ln\ N_{uft}$ is the natural logarithm of the total number of natives enrolled in university u and field of study f in academic year t; d_u, d_f and d_t are time, university and field of study fixed effects respectively; d_{tf}, d_{uf} and d_{tu} are the two-way interactions of time and field of study, university and field of study, time and university, respectively; $d_{uf} * t$ captures the interaction between university and field of study fixed effects and a time trend; ε_{uft} is an idiosyncratic error term.

Equation (2.2) encompasses the fact that some students may shift within university across fields of study due to larger inflows of foreign students. The fixed effects imply that the coefficient of interest, β_1, is identified by the variations over time within narrowly defined university - field of study cells. This should directly identify the effect of foreign students on the group of natives most closely competing with them. A non-negative estimate of β_1 implies that larger influxes of non-UK domiciled students crowd in students within universities across fields of study.

In estimating both equations (2.1) and (2.2) we face the problem that the number of foreign students enrolled is arguably endogenous to the number of native students enrolled, and Ordinary Least Squares (OLS) estimates of α_1 and β_1 would be biased. One major source of unobserved heterogeneity is represented by unobserved shocks to university available places. For instance, if universities expand and invest in building new teaching facilities and in hiring more teaching staff, they can enrol simultaneously higher levels of native and foreign students. In this case, one might find a positive spurious correlation between native and foreign numbers. Thus, in the following subsection we propose solutions for this issue.

2.5.2 Instrumental variable estimation

We use an instrumental variable strategy to address the problem of endogeneity of foreign students inflows. The ideal instrument is correlated with current flows of foreign students in universities, but uncorrelated with all the other factors that determine current flows of native students enrolled in universities.

We use the approach pioneered by Card (2001, 2005, 2009) in the labour economics literature on immigration, which uses the fact that immigrants from a particular source country tend to move into cities where migrants from their country have settled down in the past, to define an instrument to control for this potential endogeneity. The intuition is that the current flow of immigrants to a city is correlated with historical population shares into that city: a city with historically high shares of immigrants from a particular sending country is prone to receive more immigrants from that country when the national level of immigrants from the source country increases, compared to a city with historically low shares. Thus, the current inflow of immigrants from each sending country is instrumented by historical shares of immigrants into that city multiplied by the current national level of foreigners from the source country. The main assumption is that the national level inflows of foreigners from each country is exogenous to city conditions.

In the context of higher education, the conceptual analogue is that students from a particular sending country are more likely to go to universities and pursue degrees in subjects more popular among previous students from their own country. The main channel through which this prediction works is the network created among foreign students from the same country with prospective students from home. Thus, for α_1 we use the predicted flow of foreign students in university u defined as the sum over all countries of origin of the product between the share of foreigners from each country c in university u at time t_0 and the total number of foreigners from country c at time t as an instrument for the total flow of foreigners in a given university - time cell:

$$Z_{ut} = \sum_{c} \frac{F_{uct_0}}{F_{ct_0}} F_{ct} \tag{2.3}$$

where F_{uct_0} stands for the total number of foreigners domiciled in country c and enrolled in university u at time t_0; F_{ct_0} captures the total number of foreigners domiciled in country c and enrolled in all British universities at time t_0; F_{ct} is the total number inflow of foreigners domiciled in country c at current time t; time t_0 is defined as the period 1998/99-2003/04 in our estimation.[18]

[18] We use a similar instrument for β_1: the predicted flow of foreign students in university u and field of study f defined as the sum over all countries of origin of the product between the share of foreigners from each country c in university u and field of study f at time t_0 and the total number of foreigners from country c at time t as an instrument for the total flow of foreigners in a given university - field of study - time cell:

$$Z_{uft} = \sum_{c} \frac{F_{ufct_0}}{F_{ct_0}} F_{ct} \tag{2.4}$$

The relevance of the instrument rests on the notion the current relative flow of foreign students in a university is related to historical shares of foreigners in that university. In other words, enclaves of students from a specific country in a university in the past are good predictors of the current flow of students from that specific country in the university. Panel A in figure (2.3) plots the current inflow rate of foreign students in each university against the corresponding supply-push flows, while panel B plots the same measures but aggregated at university-field of study level. For reference, we have superimposed a 45-degree line on the figure. The correlation between the actual and supply-push inflows is strong, even though there are universities with lower or smaller inflows than predicted based on earlier foreign inflows. The first stage estimates presented in table (2.3) further bring evidence that our instrument is strong, satisfying the relevance criteria.

The key identification assumption is that inflows of foreign students enrolled at least six years ago are uncorrelated with other unobserved determinants of current enrolment rates of natives. Because our source of identification depends on flows of non-UK domiciled students enrolled at least six years ago, it is arguably exogenous to the sources of potential endogeneity outlined above.

2.6 Results

In this section we examine how the increase in the number of undergraduate first year foreign students enrolled in British universities affected the enrolment of native students. We proceed by first describing the overall effect, both across universities and within university - field of study pairs. We then analyse how the composition of the native student body was affected, both in terms of their their academic performance and demographic characteristics. It is worth noting that our main results are based on the group of natives enrolled directly after finishing secondary education, although we test in the next section how robust our results are when we include those natives that took gap years before pursuing an undergraduate degree.

2.6.1 Overall effect

Table (2.4) shows estimates of the effect of the increasing inflow of foreigners on natives' enrolment. Both OLS and IV estimates are presented. In the first four columns

where F_{ufct_0} is the total number of foreigners domiciled in country c and enrolled in university u and field of study f at time t_0.

data is aggregated at university-year level and the table reports estimates of α_1. In the last four columns data is aggregated at university-field of study-year level in order to account for the fact that universities have an additional margin of adjustment across fields of study; we report the estimates of β_1. In columns (1) and (3) and in columns (5) and (7) we do not control for the university specific and university-field of study specific, respectively, time variant components, whereas in columns (2) and (4) and in columns (6) and (8) we control for them. None of the four coefficients reported in columns (1)-(4) is statistically significant, suggesting that there is no effect of the influx of foreign students on natives' enrolment across universities. What emerges from comparing the OLS and the IV estimates in the last four columns, is that even though the OLS estimates seem to be positive and statistically significant, the IV estimates are not, showing that there is no evidence of either crowding in or crowding out within universities and across fields of study. We interpret these findings as resulting from the government quotas on the number of native students that was in place during the period under study.

It is worth noting that our findings are in line with the ones of Machin and Murphy (2014) who also find no evidence of any effect of overseas students on undergraduate natives enrolled in British universities, through using a similar estimation strategy to ours, but a different time period.[19]

Moreover, although various changes in the funding of higher education in the UK took place during the period under analysis, our saturated models which controls for a wide array of fixed effects ensures that our results are not affected by changes in enrolment due changes in the funding of higher education. Our results are further supported by the findings reported in Chapter 3, which show that the increases in tuition fees introduced in the UK in 2006/07 did not affect the enrolment rates of natives (see section (3.6.1) for details).

Even though we find no overall effect, there could still be distributional effects for native undergraduate students. For example, the quality of the native student body in terms of their academic performance could be altered by the larger inflows of foreign students. One hypothesis is that the marginal native student is crowded out by foreign students if universities have limited resources and recruit more able non-UK domiciled students. Another possible hypothesis is that high influxes of able foreign students increase the

[19] Machin and Murphy (2014) use an estimation in which both the dependent and the independent variable are expressed in first differences in order to account for university time varying characteristics. In our estimation, although our dependent and independent variable are expressed in levels, by including the $d_u * t$ or the $d_{uj} * t$ among the regressors we account for this.

perceived quality of the universities they attend, attracting more top performing students. Thus, in the following subsection we investigate which type of native students benefit or suffer from the higher competition from foreign students.

Given that the UK has a diverse demographic group of students, we then proceed by investigating if there are distributional effects by demographic characteristics. We mainly focus on gender, social economic status and ethnic origins.

2.6.2 Distributional effects

For the distributional analysis we adapt equations (2.1) and (2.2) in two ways. First, the outcome variable is the natural logarithm of the total number of natives with the specific characteristic j analysed. Second, we identify differentially the effect of the inflow of foreigners for each group of natives, estimating the model without a constant:

$$ln\ N_{jut} = \sum_j \gamma_j X_j + \sum_j \omega_j ln\ F_{ut} X_j + d_t + d_u + d_u * t + \varepsilon_{jut} \tag{2.5}$$

where $ln\ N_{jut}$ is the natural logarithm of natives with characteristics j enrolled in university u in academic year t (for instance, the total number of enrolled native female students) ; X_j a categorial variable equal to 1 for characteristic j.

A similar equation is estimated for the analysis of the effect across fields of study:

$$ln\ N_{juft} = \sum_j \delta_j X_j + \sum_j \sigma_j ln\ F_{ut} X_j + d_u + d_f + d_t + d_{tf} + d_{uf} + d_{tu} + d_{uf} * t + \varepsilon_{uft} \tag{2.6}$$

where $ln\ N_{juft}$ is the natural logarithm of natives with characteristics j enrolled in university u and field of study f in academic year t.

Our variables of interest are ω_j and σ_j which measure the differential effect of the inflow of foreign students on natives distribution based on characteristic j across universities and across fields of study, respectively.[20]

2.6.2.1 Native students' academic performance

We begin our analysis by focusing on the distributional effects of the native student body. We measure students' academic performance before entering university using

[20] It is worth mentioning that this type of estimation requires the data to be expanded, based on the number of categories considered for each characteristic under study.

the standardised GCSE test scores in Mathematics and English. Specifically, we use the entire cohort of students eligible for the GCSEs (both enrolled and not enrolled in university) to separate students in two groups: those who scored below and those who scored above the median grades in either English or Mathematics.

Figure (2.4) plots the number of native students who obtained a grade above the median GCSE English or in Mathematics and who are enrolled in British universities against the predicted flows of foreign students. The fitted line shows that there is strong positive correlation between the number of top native undergraduate students and the flows of mobile students. Figure (2.5) presents the correlation for the least able natives, defined as those students who took a grade below the median GCSE English or in Mathematics. The fitted line shows that although there is positive correlation between the number of less able native undergraduate students and the flows of mobile students, the correlation is much smaller than in the case of the top native students. These descriptive figures suggest that although both the less able and the best performing native students benefit, it is the latter who benefit more from the large inflows of students, rather than the marginal students.

In table (2.5) we present the results from estimating equation (2.5) and (2.6), with j referring to the distribution of the GCSE grades. In columns (1) - (4) we report estimates of ω_j and in columns (5) - (8) we present estimates of σ_j. In panel A we measure native students' ability using the English test scores. The IV estimates are statistically significant across all estimations in the university - field of study - time aggregation for top native pupils, as natives with test scores above the median tend to be crowded in. Specifically, as it can be seen in column (8) a 1% raise in the number of foreign students increases the number of native students with grades above the median GCSE English grades by 0.14%.

Panel B focuses on the Mathematics test scores and shows that marginal native students are crowded out, while the top students are crowded in by mobile students. However, the effect is statistically significant only for the best native students. Our preferred IV estimates presented in column (8) indicate that an increase of 1% in the number of foreign students increases the number of native students with grades above the median GCSE English grades by around 0.23%.

Findings so far lend support to the idea that even though on average foreign students do not affect the enrolment of natives in British universities on average, they crowd in top natives students across fields of study. We next test whether these students come from top performing secondary schools.

2.6.2.2 Quality of the secondary school attended

As the top native students benefit of the larger inflows of mobile students, we now investigate if these students come from the best secondary schools, when we measure a school's quality using the average standardised test scores in the GCSE Mathematics and English at school level.

Table (2.6) presents the estimates obtained. Panels A and B distinguish between two different measures of the quality of the school: the average GCSE English at school level and the average GCSE Mathematics at school level, respectively. Results are in line with the previous ones as it is students from top schools that are crowded in, school which are also more likely to teach the best pupils. When comparing the findings from column (8) in panels A and B, the magnitude of the effect is not very different for English and Mathematics test scores, implying that a 1% increase in the number of foreign students triggers a 0.10% increase in the number of native students coming from top secondary schools.

All in all, our results suggest that top performing students benefit from the increased competition from foreign students, and it is mainly those who attended the best state English secondary schools that are crowded in across fields of study.

2.6.2.3 Gender

We now extend the analysis to explore whether the demographic composition of the student body has been altered by the larger inflows of foreign students. We begin by exploring the distributional effects by gender given the largely documented educational gender gap (OECD (2012)).

Table (2.7) presents the estimates of the effect of the increase in foreign students on natives' gender composition. Both the OLS and the IV estimates presented in columns (1) - (4) suggest that men are crowded in by foreign students. However, our preferred IV estimates are not statistically significant and the low value of the F statistic reported in column (4) suggests that the instrument is too weak. Moreover, the IV estimates for female and male are different from each other only at the 10% significance level, as the reported p-value shows.

Columns (5) and (8) show the results obtained through exploring the variation at university-field of study level across time. Both the OLS and the IV estimates are statistically significant for men, implying a positive effect of the inflow on foreign students on native male. The IV estimates show that a 1% increase in the number of enrolled undergraduate foreigners increases the number of native males by around 0.13%. Furthermore, the estimates are different between male and female at all statistical significance levels, with a p-value of zero. The large F statistics also points to a strong instrument.

Thus, although there is no change of the distribution of natives by gender across universities due to the larger inflow of foreign students, male students are crowded-in across fields of study. We believe that these findings are in line with the ones from the analysis by academic performance when we found that top performing native students are crowding in with a larger magnitude of the effect for students performing well in Mathematics, given that among those enrolled in university native men score higher on average in GCSE Mathematics compared to women.

2.6.2.4 Social Economic Status

Given that financial constraints can be a detriment to university enrolment, in table (2.8) we present the analysis of how the higher competition from foreigners has impacted the composition of the native student body by social economic status. We use two different dimensions to define the social economic status of natives: the eligibility for free school meal at age 16 and the IDACI score.

The eligibility for free school meal is a good proxy of a student's family financial situation, as it is only students from families with various income support or benefits that are eligible for this. Panel A shows how the effect of larger foreigners enrolled differs for natives based on their free school meal eligibility. The IV estimates presented in columns (3) and (4) show no statistically significant effect for either group of natives, when we use the variation at university level. Moreover, the estimates are not statistically different from each other. When we explore variation in the number of students at university - field of study level, the IV estimates reported in columns (7) and (8) show that it is mainly students who are not eligible for free school meal who are crowded in. However, the large p-values suggest that estimates for the two groups are not statistically different from each other even at the 10% significance level.

The other measure of the social economic status that we use is the IDACI score, which is defined at local level and quantifies the proportion of children under 16 living in families that are income deprived. Thus, the larger the score, the more deprived is the area where the student was domiciled at age 16. The results presented in Panel B show that the less deprived native students are crowded in by large inflows of foreign undergraduate students, although the IV estimates are statistically significant only in columns (7) and (8). However, the estimates for the two groups are different from each other only at the 10% significance level.

We can conclude that there is limited evidence of a differential effect for natives by social economic status, with the results suggesting that if anything, it is the richer pupils that are crowded in by foreign students across fields of study.

2.6.2.5 Ethnicity

Given the wide ethnic diversity of English-born pupils, we further analyse how the distribution of natives by ethnic characteristics is affected by the inflow of foreign students. First, we distinguish between students for whom English is the main language spoken at home and those for whom it is not. Then we qualify students by their ethnicity: White, Asian, Black or other. Results are reported in table (2.9).

In panel A, the IV estimates are statistically significant across all estimations in the university - field of study - time aggregation for native pupils whose first language spoken at home is not English. As column (8) shows a 1% increase in the number of foreign students implies a 0.18% increase in the number of natives whose first language is not English pursuing an undergraduate degree. Moreover, the low p-value suggests that the estimated coefficients are statistically different by language groups. Thus, natives whose first language is not English are not crowded in across universities, but across fields of study.

Panel B brings further evidence to support this finding. Results reported in columns (1) - (8) point out that it is native students of Asian origin that seem to be crowded in by foreign students. The IV estimates are statistically significant at 1% across all estimations. That is, as columns (4) and (8) show, a 1% increase in the number of foreigners enrolled triggers an increase in the number of native students of Asian origins by 0.29 and 0.21% across universities and across fields of study, respectively. Furthermore, even though the results are not statistically significant for the other ethnic groups, the low p-value suggests that there are distributional effects across ethnic groups.

Our results suggest that English-born students whose first language is not English and who are mainly of Asian origin are crowded in across fields of studies. Moreover, UK students of Asian origin are also crowded in across universities. Findings in this subsection lend support to the idea that UK pupils from ethnic minorities, and especially of Asian origin, outperform white British pupils by the time they sit their GCSE, despite lower average attainment at earlier ages (Strand (2008), Dustmann et al. (2010), Rutter (2016), Strand (2014), Hutchinson et al. (2016)).

Our analysis of distributional effects of the foreign inflow of undergraduate students brings evidence that although there is no effect across universities, there are distributional effects within universities, across fields of study.

2.7 Robustness checks

In this section we check how robust our results are to various estimations. First, we consider all native students, even if they enrolled with gap years. Then, we also account for changes in the population of English students that could have gone to university or any behavioural changes in their university going.

2.7.1 Gap and no gap year

Our rich individual level data allows to track English pupils and to distinguish between those who enrol into university straight after secondary school and pupils who delay their entrance, through taking gap years. So far in the analysis we have differentiated between cohorts of pupils, focusing only on students who entered university without taking any gap year. However, in this section we check if our results are robust to the inclusion of native students who took gap years in our sample. The rational behind a differential effect is that native students select into taking their gap year. For instance, they could be students who are financially constraint and cannot pursue a full time undergraduate degree immediately after graduating their secondary school.

Panel A of table (2.10) presents the results of estimating equations (2.1) and (2.2), when the outcome variable includes all full time first year undergraduate natives who enrolled with or without a gap year. The findings are very similar to the ones reported in table (2.4), suggesting that those native students enrolled with gap years are not differentially affected by the inflow of mobile students, compared to native students who enrolled straight after finishing secondary school.

2.7.2 Population at risk

So far we have not accounted for the population of English pupils that could have gone to university. It could be the case that less natives are enrolled in universities due to decreasing natives population or behavioural changes in natives' university going. Thus, we further control for the population at risk among natives, accounting for those that could have enrolled as well.

We propose a way to predict the native population at risk in order to account for changes in the population of English students that could have gone to university or any behavioural changes in their university going. In theory, each individual could go to any of the 139 British universities in the choice set. However, as individuals have different characteristics the probability that one attends each of the institutions is different for each person. To operationalise the notion of population at risk, we pool the data on all English pupils who finished their compulsory education and sat their GCSEs between 2001/02-2008/09, independently of whether they enrolled in university between 2004/05-2011/12 or not, and estimate for each person, based on individual characteristics (demographics and pre-university academic performance) their probability to attend each university. Using this pooled data, we estimate a multinomial logit model of the following form:

$$Pr(y_i = u) = \frac{exp(X'\gamma_u)}{1 + \sum_u exp(X'_i\gamma_u)} \tag{2.7}$$

with $y_i = 0, 1, \ldots 139$ a categorical variable equal to 0 if native student i does not go to university or if enrolled with a gap year, and u if they go to university u; X_i are individual level characteristics which vary only across individuals and not across universities.[21] From this model we predict for each individual i their probability to enrol into each of the 139 universities and we define the population at risk for each university as the sum of these predicted probabilities.

[21] These variables are demographic characteristics (gender, ethnicity, eligibility for free school meal, special education needs indicator, IDACI score, first language spoken at home), geographical characteristics (we use the distance to the closest three universities calculated using the postcode of each of the 139 universities and the centroid of the lower layer super output area where each pupil lives at age 16 (available in the NPD), following the approach of Gibbons and Vignoles (2012) who show that geographical distance is a key factor in the university choice in England) and academic performance measures (the standardised test scores in English and in Mathematics at the GCSEs as well as the mean test scores in English and Mathematics at school level).

Thus, in panel B of table (2.10) we estimate an amended version of equation (2.2), through also controlling for the population at risk. It is worth noting that due to computational limitations, we were not able to compute the population at risk at university-field of study level. However, given that the similarity of our estimates to the one presented in columns (1) - (4) in table (2.4), we believe our results will not change in the aggregation at university - field of study level with the inclusion of this control.

To sum up, in this section we have shown that our results are not altered when we also consider natives who take gap years or when we control for changes in the university going behaviour of natives or in their cohort size.

2.8 Mechanisms

So far we have not discussed the potential mechanism that could explain our results. It could be the case that top performing native students are crowded in by foreign students as the latter bring additional financial resources to universities. In particular, if this was the case we would expect the effect of the inflow of foreign students domiciled outside the EU to be positive as they are paying much larger tuition fees. We test for this hypothesis by distinguishing between foreign students by their domicile. Moreover, we further test this hypothesis by distinguishing between Russell Group and Non-Russell Group universities as the former group includes universities with high international reputation which attract large inflows of international students and also charge higher tuition fees compared to the rest.

Apart from financial resources, foreign student could also crowd in top performing natives if they increase the quality of the university attended. This could happen if, for instance, non-UK domiciled students are very able students. We test for this hypothesis by investigating the effect of the inflow of foreign students on university ranking.

2.8.1 EU vs Non-EU students

Given that universities distinguish between EU and non-EU students in terms of the level of tuition fees and the regulations for places available and financial support, we further analyse the robustness of our results when the inflow of foreigners is differentiated between EU and non-EU domiciled students. In panel A of table (2.10) we measure the inflow of foreign students only as the inflow of non-EU domiciled students, while in panel B we focus only on EU students rather than the total foreigners as we have done so far. In each panel we present estimates of α_1 and β_1 from equations (2.1) and (2.2) when we use different measures of the mobile students.

What emerges from the comparison across columns and these two panels, is that in the IV estimations, which are our preferred estimations, the overall effect is not statistically significant for either measure of foreign inflows. When comparing these results to the ones reported in table (2.4) the estimates are quite similar, suggesting that our findings are not driven by the higher financial resources brought by non-UK domiciled students.

2.8.2 Russell Group vs. Non-Russell Group university

We proceed by distinguishing between Russell Group and Non-Russell Group universities as the former group includes universities with high international reputation which attract large inflows of international students and also charge higher tuition fees compared to the rest. For this we divide universities based on their belonging to the Russell Group which includes the best 20 research universities in the UK, which were part of the group until 2011/12. Results are reported in panel C of table (2.11). Columns (1) and (2) present the estimates of ω_j from estimating (2.5). The IV results show a crowding out from top universities due to larger inflows of foreign students. Our preferred estimate, reported in column (2), is significant only at the 10% significance level. Moreover, the low F statistic, which is considerably below 10, indicates that the results should be interpreted with caution as the instrument is weak.

In columns (3) and (4) the analysis is done using data grouped at university-field of study and estimates of σ_j are reported. The large F statistics suggests a strong instrument. The positive estimates presented in column (4) for the non-Russell group and the negative effect for the Russell group suggest that our baseline estimates are not triggered by the larger fees paid by foreign students enrolled in top universities, although none of the estimates is statistically significant.

2.8.3 University ranking

We proceed by exploring another potential mechanism. In particular, we want to understand whether the large inflow of foreign students crowds in top performing natives as they had a positive effect on the ranking of the university attended. For this we use an additional data set, called the Sunday Times Good University Guide between

2004/05-2011/12.[22] This league table is published yearly and ranks around 120 British universities each year. As discussed in section 2.3 the importance of British league tables to prospective students has been documented in the literature, with students sorting into universities based on these rankings. We use the overall university ranking which is derived using a comprehensive list of scores including expenditure per student, student-staff ratio, job prospects, university entry scores, teaching or research quality. The ranking represents a comparable index of university quality, allowing us to further bring light on whether universities are becoming more competitive due to the higher inflow of foreign students.

Panel D in table (2.11) shows the estimates of α_1 from estimating equation (2.1) when the outcome variable is the overall ranking of university. As we have data only at university level, we cannot run the analysis at university-field of study level. Both the OLS and the IV estimates presented in columns (1) and (2), respectively, are positive. The latter suggests that a 1% increase in the number of foreign students triggered an increase in the university ranking of around 0.07. However, the estimate is only significant at the 10% significance level and the F statistics is just below 10. Thus, we interpret this result as suggestive that the influx of foreign students increased the quality of universities, but the effect is of small magnitude and only just significant.

Our analysis indicates that the crowding in effect we identify for top performing natives is not due to the larger financial resources brought by non-EU domiciled students. We find limited evidence that the inflow of mobile students triggered a slight increase in the quality of universities.

2.9 Discussion and conclusion

In this paper we study whether the inflows of first year full-time undergraduate foreign students enrolled in British universities had any impact on the enrolment of native undergraduate students. We combine very rich individual level administrative data on eight cohorts of English students and on non-UK domiciled undergraduate students to run the analysis. We employ an IV estimation in order to control for the potential endogeneity of the influxes of mobile students. Specifically, we use historical shares of students from a sending country enrolled into a university together with current national changes in the stock of students from this country to instrument the current flows of foreign undergraduate students attending a university. By using a rich array of fixed

[22] The data was kindly provided by Alastair McCall, editor of the The Sunday Times Good University Guide.

effects we also ensure that the various of changes in the higher education funding that took place during the period under analysis did not affect our results.

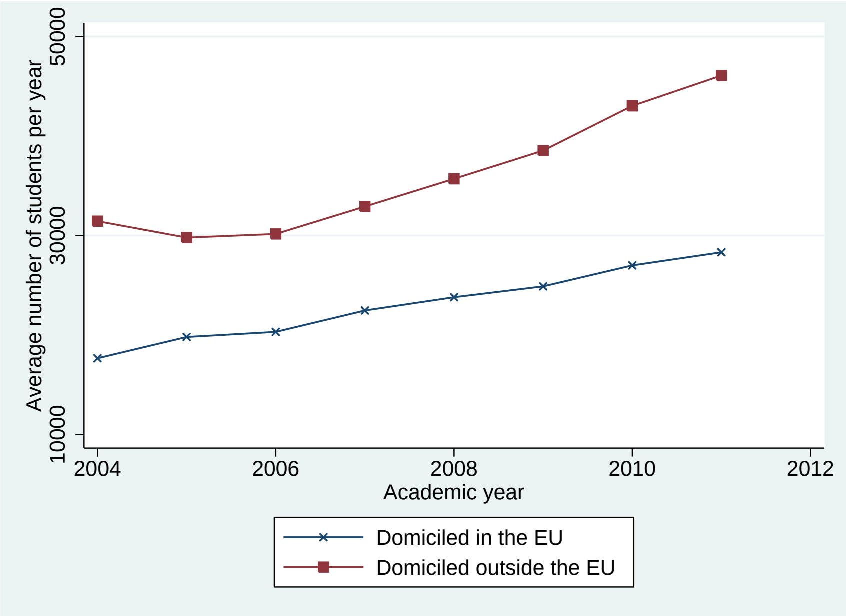
Our results confirm previous findings in the literature that there are no effects on average. As these overall effects could mask distributional effects we extend the analysis to offer the first analysis of changes in the composition of enrolled native undergraduates due to the larger number of enrolled foreign students in British universities. Our results show that it is mainly the top performing native students and English pupils from top secondary schools that benefit from the increased enrolment rates of foreign students in British universities. Moreover, we find that male natives and natives whose first language is not English, as well as natives of Asian ethnic origins are crowded in by foreign students. Our distributional analysis by natives' demographic composition complements our distributional analysis by natives' ability, as there is a large literature that shows that in England students of Asian origins tend to score better than UK-born white pupils in the GCSE.

Given that all our results are identified within universities across fields of study, we believe that our findings support the idea that some students shift within university, across fields of study due to the larger inflows of foreign students. Our results suggest that universities benefit from enrolling more foreign students as they seem to attract more able native students, becoming more competitive. From an equity point of view, universities also attract more natives from minority groups which were under-represented in higher education.

Our analysis of potential mechanisms shows that the crowding in effect we identify for top performing natives is not due to the larger financial resources brought by students domiciled outside the EU who pay on average much higher tuition fees compared to natives and EU-domiciled students. We find limited evidence that the inflow of foreign students crowds in top natives through increasing the quality of the university. Potential alternative mechanisms for our distributional effects could be that the high quality foreign students enrolled trigger also an increase in the perceived quality of the universities they attend and consequently attract more able native students. Moreover, it could be the case that top performing students prefer to enhance their student experience by enrolling into a university with a culturally rich international environment.

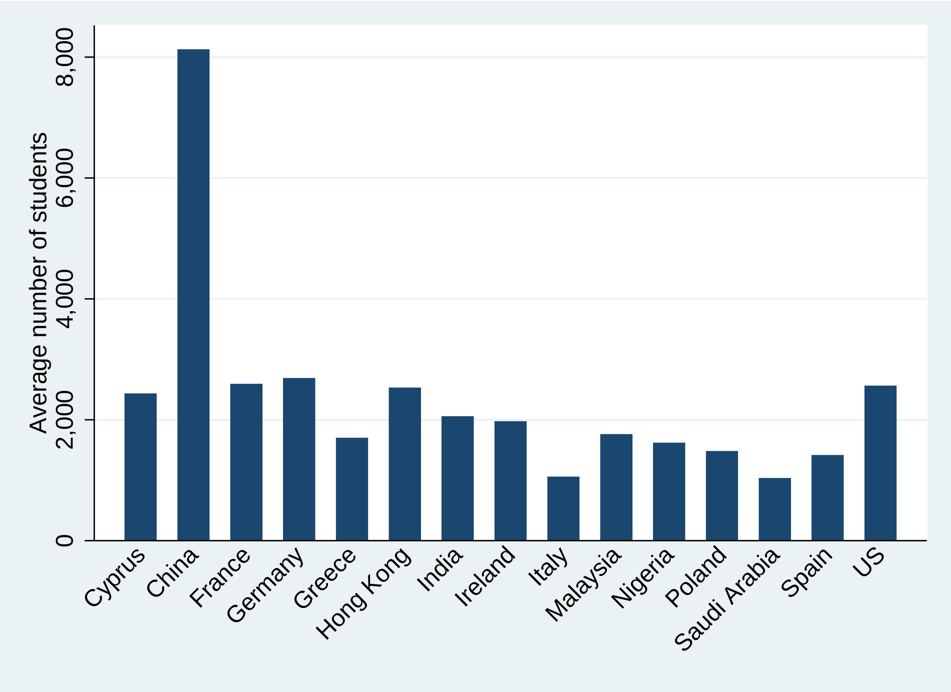
Tables and figures

FIGURE 2.1: Flows of foreign students by domicile



Notes: The figure depicts the average number of foreign undergraduate students enrolled in English universities between 2004/05-2011/12, distinguishing between those domiciled in the EU and those domiciled outside the EU. Source: Author's calculations using HESA student record data.

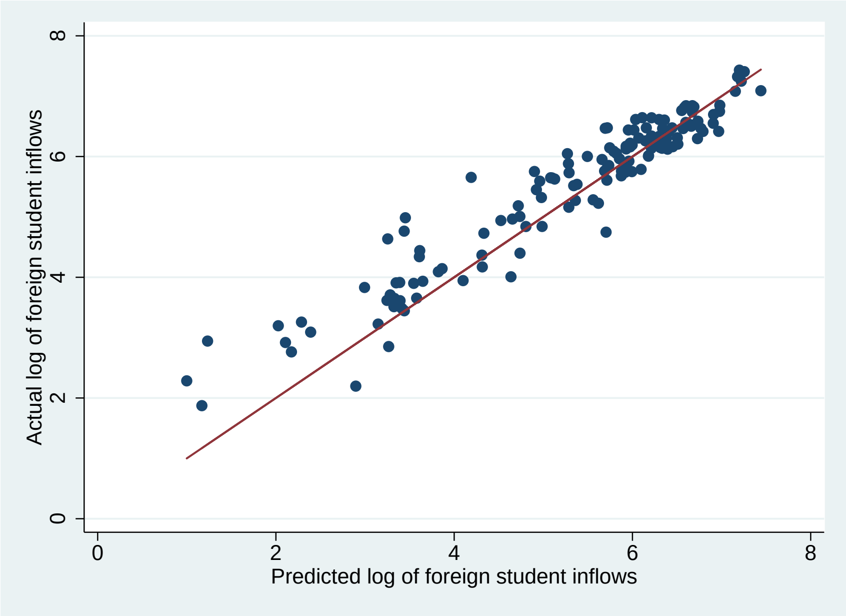
FIGURE 2.2: Top nationality among full-time first year undergraduate students



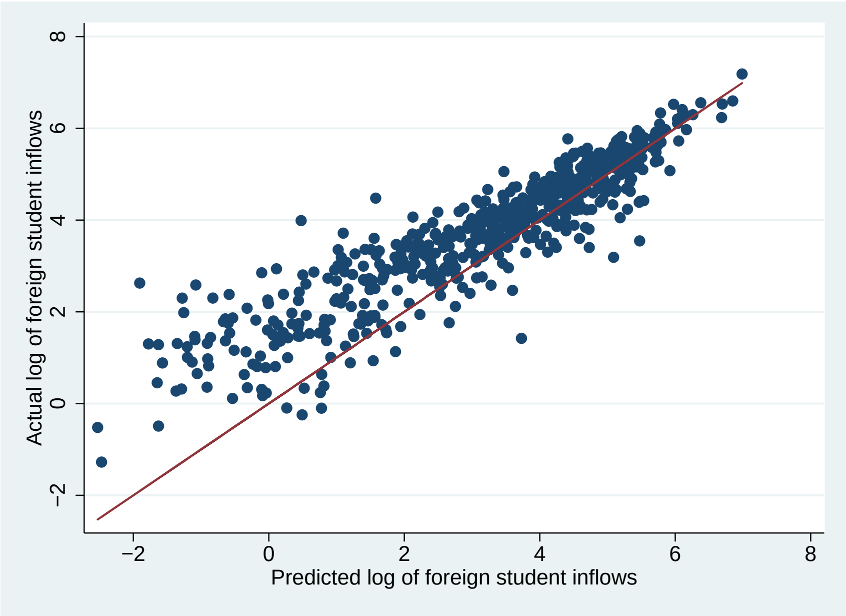
Notes: The figure depicts the top 15 nationalities among foreign undergraduate students enrolled in English universities between 2004/05-2011/12. Source: Author's calculations using HESA student record data.

FIGURE 2.3: Actual and supply-driven inflows of foreign students

Panel A: University level



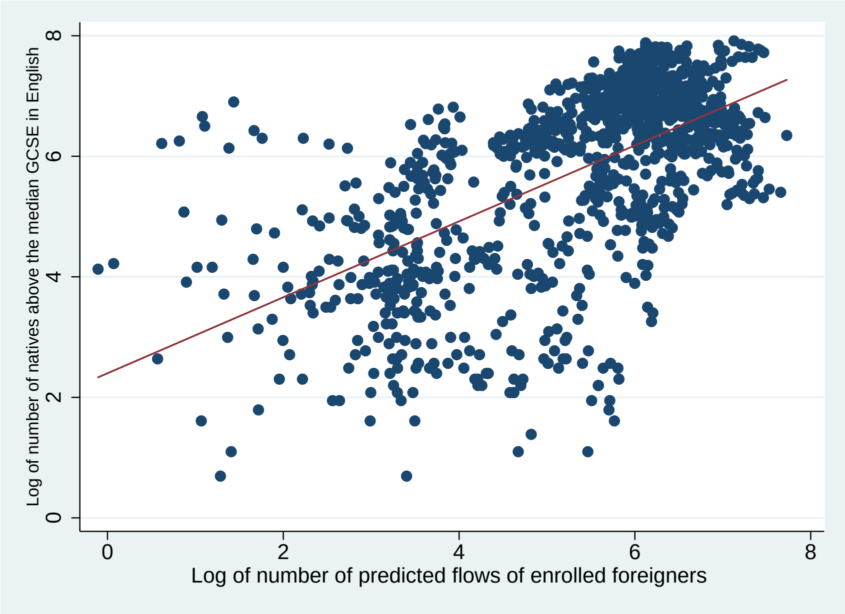
Panel B: University-field of study level



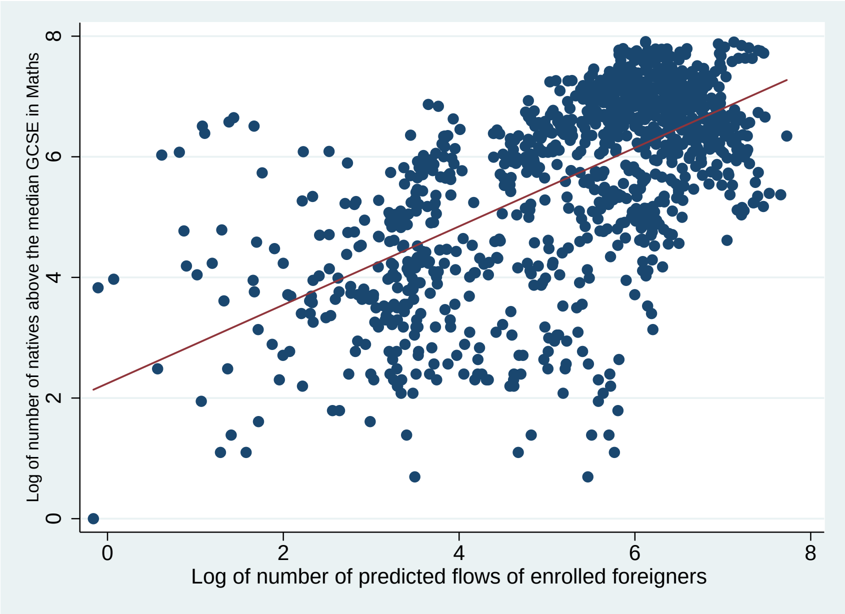
Notes: The graph in panel A plots the actual inflow of foreign undergraduate students enrolled in English universities between 2004/05-2011/12 against the predicted inflows; the aggregation is at university level. The graph in panel A plots the actual inflow of foreign undergraduate students enrolled in English universities between 2004/05-2011/12 against the predicted inflows; the aggregation is at university - field of study level. Source: Author's calculations using linked NPD-HESA and the HESA student record data.

FIGURE 2.4: Most able natives and predicted flows of foreign students enrolled

Panel A: GCSE English



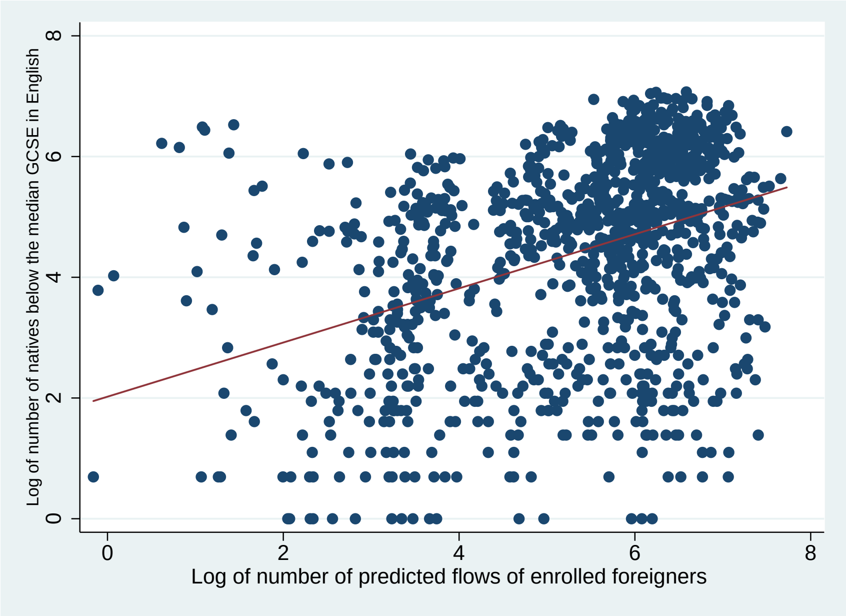
Panel B: GCSE Mathematics



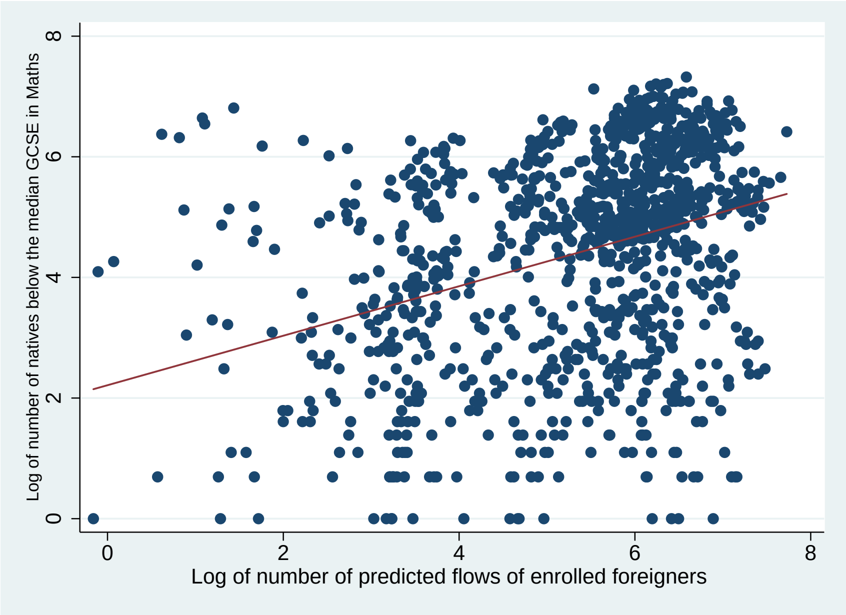
Notes: The graph in panel A plots the correlation between the inflow of students who scored above the median in GCSE English against the predicted flows of enrolled foreign students. The graph in panel A plots the correlation between the inflow of students who scored above the median in GCSE Mathematics against the predicted flows of enrolled foreign students. Source: Author's calculations using linked NPD-HESA and the HESA student record data.

FIGURE 2.5: Least able natives and predicted flows of foreign students enrolled

Panel A: GCSE English



Panel B: GCSE Mathematics



Notes: The graph in panel A plots the correlation between the inflow of students who scored below the median in GCSE English against the predicted flows of enrolled foreign students. The graph in panel A plots the correlation between the inflow of students who scored below the median in GCSE Mathematics against the predicted flows of enrolled foreign students. Source: Author's calculations using linked NPD-HESA and the HESA student record data.

TABLE 2.1: University related characteristics by domicile

	Mean	SD	N
Panel A: Students domiciled in the UK			
Enrolled in a Russell group university	0.188	0.390	1,426,587
University by location			
English University	0.959	0.199	1,426,587
Welsh University	0.029	0.169	1,426,587
Scottish University	0.0110	0.104	1,426,587
Northern Ireland University	0.001	0.027	1,426,587
Field of study			
Medicine, Dentistry and Allied Subjects	0.226	0.418	1,426,587
STEM	0.193	0.394	1,426,587
Social Sciences	0.294	0.456	1,426,587
Languages and History	0.114	0.317	1,426,587
Arts, Education, Other	0.173	0.378	1,426,587
Panel B: Students domiciled outside the UK			
Domiciled in an EU country	0.390	0.488	471,935
Domiciled in a non-EU country	0.610	0.488	471,935
Enrolled in a Russell group university	0.280	0.449	471,935
University by location			
English University	0.853	0.354	471,935
Welsh University	0.043	0.202	471,935
Scottish University	0.091	0.288	471,935
Northern Ireland University	0.013	0.113	471,935
Field of study			
Medicine, Dentistry and Allied Subjects	0.130	0.337	471,935
STEM	0.227	0.419	471,935
Social Sciences	0.415	0.493	471,935
Languages and History	0.101	0.301	471,935
Arts, Education, Other	0.126	0.332	471,935

Notes: The table shows university specific characteristics for all foreign 1st year undergraduate students enrolled in British universities domiciled in EU and outside the EU and for English pupils who sat the GCSEs in English state school and who enrolled as first year full time undergraduate in a British university at age 18/19 between 2004/05-2011/12. All reported variables are categorial variables equal to 1 for the specific variable. Source- author's own calculations using the NPD-HESA linked data.

TABLE 2.2: Demographic characteristics and academic performance of natives

	Mean	SD	N
Panel A: Demographics			
Female	0.549	0.498	1,426,587
White	0.802	0.398	1,426,587
Black	0.049	0.215	1,426,587
Asian	0.112	0.315	1,426,587
Other	0.037	0.189	1,426,587
English as first language	0.862	0.345	1,426,587
Free school meal	0.073	0.260	1,426,587
IDACI score	0.170	0.161	1,426,587
Panel B: GCSE academic performance			
No subjects sat GCSE	10.230	1.417	1,426,587
At least 5 A*-C GCSE	0.881	0.324	1,426,587
At least 5 A*-G GCSE	0.994	0.077	1,426,587
Std GCSE English	0.704	0.665	1,426,587
Std GCSE Mathematics	0.692	0.699	1,426,587
Average std GCSE English at school level	0.175	0.462	1,426,587
Average std GCSE Mathematics at school level	0.175	0.472	1,426,587

Notes: The table shows demographic and secondary school characteristics measured at age 16 for all English pupils who sat the GCSEs in English state school and who enrolled as first year full time undergraduate in a British university at age 18/19 between 2004/05-2011/12. The IDACI score is a number between 0 and 1: the higher it is, the worse off the children are. Source: author's own calculations using the NPD-HESA linked data.

TABLE 2.3: First stage estimates

	University level		University-field level	
	(1)	(2)	(3)	(4)
ln predicted Foreigners	0.610^{***}	0.569^{***}	0.416^{***}	0.364^{***}
	(0.112)	(0.161)	(0.039)	(0.042)
F statistic	29.423	12.439	115.906	75.458
Universities	139	139	139	139
Observations	1,112	1,112	4,915	4,915
University FE	X	X	X	X
Time FE	X	X	X	X
Field of study FE			X	X
University FE X Time FE			X	X
Field of study FE X Time FE			X	X
University FE X Field of study FE			X	X
University FE X Time trend		X		
University FE X Field of study FE X Time trend				X

Notes: The regressions in columns (1) - (2) use data on 139 universities, observed over 8 years. The regressions in columns (3) - (4) use data on 139 universities grouped in 5 fields of study, observed over 8 years. The outcome variable is the logarithm of the total number of English students who sat their GCSE in a state secondary school between 2001/02-2008/09 and who enrolled as first year undergraduates in each university without a gap year, between 2004/05-2011/12. Robust standard errors clustered at university level in parentheses. F statistics is based on the Kleinbergen-Paap Wald F statistics. *** p<0.01, ** p<0.05, * p<0.1

TABLE 2.4: Overall effect of foreign students on natives' enrolment

	University level				University-field level			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln Foreigners	-0.023	0.016	-0.002	-0.093	0.071***	0.040**	0.060	0.039
	(0.046)	(0.058)	(0.136)	(0.152)	(0.021)	(0.017)	(0.043)	(0.045)
F statistic			29.423	12.439			115.906	75.458
Universities	139	139	139	139	139	139	139	139
Observations	1,112	1,112	1,112	1,112	4,915	4,915	4,915	4,915
University FE	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X
Field of study FE					X	X	X	X
University FE X Time FE					X	X	X	X
Field of study FE X Time FE					X	X	X	X
University FE X Field of study FE					X	X	X	X
University FE X Time trend		X		X				
University FE X Field of study FE X Time trend						X		X

Notes: The regressions in columns (1)-(4) use data on 139 universities, observed over 8 years. The regressions in columns (5)-(8)use data on 139 universities grouped in 5 fields of study, observed over 8 years. The outcome variable is the total number of English students who sat their GCSE in a state secondary school between 2001/02-2008/09 and who enrolled as first year undergraduates in each university between 2004/05-2011/12. Robust standard errors clustered at university level in parentheses. F statistics is based on the Kleinbergen-Paap Wald F statistics. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 2.5: Effect on natives' enrolment by students' ability

	University level				University-field level			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: GCSE English								
ln Foreigners X Below median	-0.061	0.026	0.021	0.085	-0.009	-0.033	-0.002	-0.007
	(0.067)	(0.072)	(0.151)	(0.145)	(0.026)	(0.025)	(0.044)	(0.048)
ln Foreigners X Above median	0.113*	0.200***	0.177	0.241**	0.142***	0.119***	0.147***	0.142***
	(0.061)	(0.065)	(0.124)	(0.122)	(0.026)	(0.024)	(0.042)	(0.046)
F statistic			16.368	9.191			58.143	40.681
P-value	0.022	0.027	0.076	0.086	0.001	0.001	0.002	0.003
Universities	139	139	139	139	139	139	139	139
Observations	2,224	2,224	2,224	2,224	9,830	9,830	9,830	9,830
Panel B: GCSE Mathematics								
ln Foreigners X Below median	-0.136**	-0.087	-0.052	-0.049	-0.051**	-0.071***	-0.042	-0.044
	(0.062)	(0.074)	(0.156)	(0.165)	(0.025)	(0.024)	(0.046)	(0.048)
ln Foreigners X Above median	0.123**	0.172**	0.191	0.194	0.202***	0.182***	0.226***	0.225***
	(0.061)	(0.071)	(0.145)	(0.158)	(0.026)	(0.023)	(0.046)	(0.046)
F statistic			16.368	9.191			58.143	40.681
P-value	0.001	0.001	0.020	0.024	0.000	0.000	0.000	0.000
Universities	139	139	139	139	139	139	139	139
Observations	2,224	2,224	2,224	2,224	9,830	9,830	9,830	9,830
University FE	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X
Field of study FE					X	X	X	X
University FE X Time FE					X	X	X	X
Field of study FE X Time FE					X	X	X	X
University FE X Field of study FE					X	X	X	X
University FE X Time trend		X		X				
University FE X Field of study FE X Time trend						X		X

Notes: The regressions in columns (1)-(4) use data on 139 universities, observed over 8 years. The regressions in columns (5)-(8)use data on 139 universities grouped in 5 fields of study, observed over 8 years. The outcome variable is the total number of English students who sat their GCSE in a state secondary school between 2001/02-2008/09 and who enrolled as first year undergraduates in each university between 2004/05-2011/12 separated by their GCSE English test scores and by their GCSE Mathematics test scores in panels A and B, respectively. All regressions in panel A control for the English test score dummy: below the median GCSE English and above the median test score in GCSE English as the regressions are estimated without a constant. All regressions in panel B control for the Mathematics test score dummy: below the median GCSE Mathematics and above the median test score in GCSE Mathematics as the regressions are estimated without a constant. Robust standard errors clustered at university level in parentheses. F statistics is based on the Kleinbergen-Paap Wald F statistics. The P-value corresponds to the F test of estimates presented in each panel being equal to each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 2.6: Effect on natives' enrolment by the quality of the attended secondary school

	University level				University-field level			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: GCSE English								
ln Foreigners X Below median	0.035	0.109*	0.073	0.023	0.044**	0.018	0.049	0.037
	(0.059)	(0.063)	(0.130)	(0.123)	(0.020)	(0.018)	(0.040)	(0.039)
ln Foreigners X Above median	0.039	0.114**	0.055	0.006	0.103***	0.077***	0.104**	0.091**
	(0.052)	(0.057)	(0.122)	(0.117)	(0.020)	(0.016)	(0.042)	(0.038)
F statistic			16.368	9.191			58.143	40.669
P-value	0.911	0.914	0.692	0.701	0.003	0.004	0.020	0.024
Universities	139	139	139	139	139	139	139	139
Observations	2,224	2,224	2,224	2,224	9,830	9,830	9,830	9,830
Panel B: GCSE Mathematics								
ln Foreigners X Below median	0.025	0.093	0.068	0.011	0.035*	0.010	0.043	0.035
	(0.061)	(0.061)	(0.141)	(0.142)	(0.021)	(0.018)	(0.039)	(0.038)
ln Foreigners X Above median	0.045	0.113**	0.067	0.010	0.105***	0.080***	0.109***	0.101***
	(0.054)	(0.055)	(0.133)	(0.136)	(0.020)	(0.016)	(0.040)	(0.036)
F statistic			16.368	9.191			58.143	40.669
P-value	0.640	0.651	0.982	0.983	0.001	0.001	0.004	0.006
Universities	139	139	139	139	139	139	139	139
Observations	2,224	2,224	2,224	2,224	9,830	9,830	9,830	9,830
University FE	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X
Field of study FE					X	X	X	X
University FE X Time FE					X	X	X	X
Field of study FE X Time FE					X	X	X	X
University FE X Field of study FE					X	X	X	X
University FE X Time trend		X		X				
University FE X Field of study FE X Time trend						X		X

Notes: The regressions presented in columns (1)-(4) use data on 139 universities, observed over 8 years. The regressions in columns (5)-(8)use data on 139 universities grouped in 5 fields of study, observed over 8 years. The outcome variable is the total number of English students who sat their GCSE in a state secondary school between 2001/02-2008/09 and who enrolled as first year undergraduates in each university between 2004/05-2011/12 separated by their secondary school average GCSE English test scores and by their GCSE Mathematics test scores in panels A and B, respectively. All regressions in panel A control for the English test score dummy: below the median average GCSE English test score at school level and above the median average GCSE English test score at school level as the regressions are estimated without a constant. All regressions in panel B control for the Mathematics test score dummy: below the average GCSE Mathematics test score at school level and above average GCSE Mathematics test score at school level as the regressions are estimated without a constant. Robust standard errors clustered at university level in parentheses. F statistics is based on the Kleinbergen-Paap Wald F statistics. The P-value corresponds to the F test of estimates presented in each panel being equal to each other. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 2.7: Effect on natives' enrolment by gender

	University level				University-field level			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln Foreigners X Male	0.059	0.126**	0.102	0.027	0.117***	0.091***	0.126***	0.126***
	(0.050)	(0.054)	(0.129)	(0.130)	(0.017)	(0.016)	(0.037)	(0.037)
ln Foreigners X Female	-0.001	0.066	0.054	-0.022	0.017	-0.010	0.009	0.009
	(0.054)	(0.056)	(0.129)	(0.132)	(0.019)	(0.018)	(0.035)	(0.034)
F statistic			16.368	9.191			58.143	40.681
P-value	0.029	0.035	0.074	0.083	0.000	0.000	0.000	0.000
Universities	139	139	139	139	139	139	139	139
Observations	2,224	2,224	2,224	2,224	9,830	9,830	9,830	9,830
University FE	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X
Field of study FE					X	X	X	X
University FE X Time FE					X	X	X	X
Field of study FE X Time FE					X	X	X	X
University FE X Field of study FE					X	X	X	X
University FE X Time trend		X		X				
University FE X Field of study FE X Time trend						X		X

Notes: The regressions in columns (1)-(4) use data on 139 universities, observed over 8 years. The regressions in columns (5)-(8)use data on 139 universities grouped in 5 fields of study, observed over 8 years. The outcome variable is the total number of English students who sat their GCSE in a state secondary school between 2001/02-2008/09 and who enrolled as first year undergraduates in each university between 2004/05-2011/12 separated by gender. All regressions control for the gender dummies (female and male) as the model is estimated without a constant. Robust standard errors clustered at university level in parentheses. F statistics is based on the Kleinbergen-Paap Wald F statistics. The P-value corresponds to the F test of estimates presented in each panel being equal to each other. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 2.8: Effect on natives' enrolment by social economic status

	University level				University-field level			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Free school meal eligibility								
ln Foreigners X Eligible for free school meal	0.084	0.166**	0.122	0.136	0.026	0.007	0.044	0.054
	(0.058)	(0.064)	(0.117)	(0.109)	(0.023)	(0.022)	(0.043)	(0.042)
ln Foreigners X Non-Eligible for free school meal	0.070	0.151**	0.082	0.096	0.122***	0.103***	0.109***	0.120***
	(0.052)	(0.066)	(0.125)	(0.108)	(0.024)	(0.023)	(0.041)	(0.039)
F statistic			16.368	9.191			58.143	40.669
P-value	0.803	0.809	0.502	0.516	0.007	0.008	0.119	0.131
Universities	139	139	139	139	139	139	139	139
Observations	2,224	2,224	2,224	2,224	9,830	9,830	9,830	9,830
Panel B: IDACI score								
ln Foreigners X Below median IDACI	-0.002	0.058	0.041	-0.025	0.070***	0.043**	0.068	0.061
	(0.052)	(0.055)	(0.128)	(0.127)	(0.021)	(0.018)	(0.042)	(0.039)
ln Foreigners X Above median IDACI	0.067	0.127**	0.136	0.070	0.089***	0.062***	0.111***	0.104**
	(0.054)	(0.057)	(0.127)	(0.129)	(0.022)	(0.020)	(0.041)	(0.040)
F statistic			16.368	9.191			58.143	40.681
P-value	0.133	0.146	0.041	0.048	0.404	0.419	0.085	0.095
Universities	139	139	139	139	139	139	139	139
Observations	2,224	2,224	2,224	2,224	9,830	9,830	9,830	9,830
University FE	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X
Field of study FE					X	X	X	X
University FE X Time FE					X	X	X	X
Field of study FE X Time FE					X	X	X	X
University FE X Field of study FE					X	X	X	X
University FE X Time trend		X		X				
University FE X Field of study FE X Time trend						X		X

Notes: The regressions presented in columns (1)-(4) use data on 139 universities, observed over 8 years. The regressions presented in columns (5)-(8) use data on 139 universities grouped in 5 fields of study, observed over 8 years. The outcome variable is the total number of English students who sat their GCSE in a state secondary school between 2001/02-2008/09 and who enrolled as first year undergraduates in each university between 2004/05-2011/12 separated by their free school meal eligibility and by their IDACI score in panels A and B, respectively. All regressions control for the language dummies (eligible for free school meal and not eligible for free school meal) or the IDACI score(below the median of the IDACI score and above the median of the IDACI score) in panel A and B, respectively as they are estimated without a constant. Robust standard errors clustered at university level in parentheses. F statistics is based on the Kleinbergen-Paap Wald F statistics. The P-value corresponds to the F test of estimates presented in each panel being equal to each other. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 2.9: Effect on natives' enrolment by ethnic origins

	University level				University-field level			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: First language spoken at home								
ln Foreigners X English not 1st language	0.248***	0.262***	0.286**	0.214	0.147***	0.119***	0.180***	0.184***
	(0.073)	(0.073)	(0.138)	(0.168)	(0.027)	(0.025)	(0.047)	(0.046)
ln Foreigners X English 1st language	-0.114	-0.100	-0.135	-0.207	-0.005	-0.033	-0.045	-0.040
	(0.073)	(0.086)	(0.164)	(0.187)	(0.027)	(0.025)	(0.047)	(0.044)
F statistic			16.368	9.191			58.143	40.671
P-value	0.001	0.001	0.000	0.000	0.001	0.001	0.000	0.000
Universities	139	139	139	139	139	139	139	139
Observations	2,224	2,224	2,224	2,224	9,830	9,830	9,830	9,830
Panel B: Ethnic group								
ln Foreigners X Asian	0.304***	0.303***	0.275**	0.299***	0.173***	0.158***	0.207***	0.213***
	(0.070)	(0.061)	(0.131)	(0.112)	(0.022)	(0.021)	(0.031)	(0.032)
ln Foreigners X White	-0.077	-0.078	-0.171	-0.148	0.031	0.015	-0.007	-0.001
	(0.080)	(0.084)	(0.153)	(0.128)	(0.034)	(0.034)	(0.047)	(0.046)
ln Foreigners X Black	0.063	0.063	0.006	0.030	-0.008	-0.024	-0.004	0.003
	(0.059)	(0.055)	(0.121)	(0.113)	(0.022)	(0.021)	(0.034)	(0.034)
ln Foreigners X Other	0.049	0.048	-0.021	0.002	0.000	-0.015	-0.005	0.001
	(0.054)	(0.048)	(0.115)	(0.103)	(0.017)	(0.015)	(0.029)	(0.030)
F statistic			8.197	4.758			29.133	21.071
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Universities	139	139	139	139	139	139	139	139
Observations	4,448	4,448	4,448	4,448	19,660	19,660	19,660	19,660
University FE	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X
Field of study FE					X	X	X	X
University FE X Time FE					X	X	X	X
Field of study FE X Time FE					X	X	X	X
University FE X Field of study FE					X	X	X	X
University FE X Time trend		X		X				
University FE X Field of study FE X Time trend						X		X

Notes: The regressions in columns (1)-(4) use data on 139 universities, observed over 8 years. The regressions in columns (5)-(8)use data on 139 universities grouped in 5 fields of study, observed over 8 years. The outcome variable is the total number of English students who sat their GCSE in a state secondary school between 2001/02-2008/09 and who enrolled as first year undergraduates in each university between 2004/05-2011/12 separated by first language spoken at home and by their ethnic group in panels A and B, respectively. All regressions in panel A control for the language dummies: English as first language spoken at home and English not the first language spoken at home as the regressions are estimated without a constant. All regressions in panel B control for ethnicity dummies: White, Asian, Black and other as the regressions are estimated without a constant. Robust standard errors clustered at university level in parentheses. F statistics is based on the Kleinbergen-Paap Wald F statistics. The P-value corresponds to the F test of estimates presented in each panel being equal to each other. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 2.10: Robustness checks

	University level				University-field level			
	OLS	OLS	IV	IV	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Gap and no gap years								
ln Foreigners	-0.035	0.008	0.019	-0.137	0.052***	0.046***	0.040	0.044
	(0.036)	(0.042)	(0.121)	(0.134)	(0.018)	(0.015)	(0.036)	(0.029)
F statistic			29.397	12.439			115.788	75.458
Universities	139	139	139	139	139	139	139	139
Observations	1,112	1,112	1,112	1,112	4,915	4,915	4,915	4,915
Panel B: Population at risk								
ln Foreigners	-0.026	0.014	0.008	-0.103				
	(0.046)	(0.058)	(0.132)	(0.152)				
ln Predicted Population	-0.392**	-0.292*	-0.381**	-0.343**				
	(0.160)	(0.163)	(0.181)	(0.167)				
F statistic			30.353	12.241				
Universities	139	139	139	139				
Observations	1,112	1,112	1,112	1,112				
University FE	X	X	X	X	X	X	X	X
Time FE	X	X	X	X	X	X	X	X
Field of study FE					X	X	X	X
University FE X Time FE					X	X	X	X
Field of study FE X Time FE					X	X	X	X
University FE X Field of study FE					X	X	X	X
University FE X Time trend		X		X				
University FE X Field of study FE X Time trend						X		X

Notes: The regressions reported in columns (1)-(4) use data on 139 universities, observed over 8 years. The regressions presented in columns (5)-(8) use data on 139 universities grouped in 5 fields of study, observed over 8 years. In panel A, the outcome variable is the total number of English students who sat their GCSE in a state secondary school enrolled in each university between 2004/05-2011/12 with or without taking gap years. In panel B, the outcome variable is the total number of English students who sat their GCSE in a state secondary school between 2001/02-2008/09 and who enrolled as first year undergraduates in each university between 2004/05-2011/12. Robust standard errors clustered at university level in parentheses.*** p<0.01, ** p<0.05, * p<0.1.

TABLE 2.11: Mechanisms

	University level		University-field level	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Panel A: Non-EU students				
ln Non-EU	-0.003	-0.060	0.012	-0.001
	(0.038)	(0.087)	(0.014)	(0.042)
F statistic		13.950		58.297
Universities	139	139	139	139
Observations	1,112	1,112	4,915	4,915
Panel B: EU students				
ln EU	0.044	0.134	0.056***	0.038
	(0.040)	(0.082)	(0.016)	(0.047)
F statistic		8.638		79.612
Universities	139	139	139	139
Observations	1,112	1,112	4,915	4,915
Panel C: Russell Group vs. Non-Russell Group				
ln Foreigners in Non-RG universities	0.017	-0.089	0.043**	0.050
	(0.061)	(0.153)	(0.019)	(0.051)
ln Foreigners in RG universities	-0.000	-0.274*	0.017	-0.025
	(0.088)	(0.165)	(0.033)	(0.078)
F statistic		6.575		33.948
P-value	0.870	0.237	0.499	0.424
Universities	139	139	139	139
Observations	1,112	1,112	4,915	4,915
Panel D: University ranking				
ln Foreigners	1.564	6.656*		
	(1.355)	(3.726)		
F statistic		9.646		
Universities	122	122		
Observations	945	945		
University FE	X	X	X	X
Time FE	X	X	X	X
Field of study FE			X	X
University FE X Time FE			X	X
Field of study FE X Time FE			X	X
University FE X Field of study FE			X	X
University FE X Time trend	X	X		
University FE X Field of study FE X Time trend			X	X

Notes: The regressions reported in columns (1)-(4) use data on 139 universities, observed over 8 years. The regressions presented in columns (5)-(8) use data on 139 universities grouped in 5 fields of study, observed over 8 years. The outcome variable each of the three panels is the total number of English students who sat their GCSE in a state secondary school enrolled in each university between 2004/05-2011/12. In panel A, we measure the inflow of foreign students as the total number of non-EU domiciled students enrolled. In panel B, we measure the inflow of foreign students as the total number of EU domiciled students enrolled. In panel C, we distinguish between Russell and Non-Russell group universities. In Panel D, the outcome variable is the rank of the university. Robust standard errors clustered at university level in parentheses.*** p<0.01, ** p<0.05, * p<0.1.

Chapter 3

Higher education funding and early geographic mobility

joint with Ghazala Azmat

3.1 Introduction

Over the last 20 years, tertiary education funding has changed considerably around the world. According to OECD data, since 1995, 14 out of the 25 OECD countries for which data is available have reformed their university funding schemes, leading mainly to increases in tuition fees, as well as changes in the level of subsidies available to students.[1]

In England and Wales, since the introduction of tuition fees in the academic year 1998/99, various changes in the funding of higher education have taken place. Starting from 1998/99 students were required to pay up to £1,000 per year for tuition, where a typical degree would cover three years. In conjunction with changes in tuition fees, maintenance grants were first halved in 1998 and then abolished in 1999/00, while maintenance loans were increased in 1998/99 in an attempt to offset these adverse changes. In 2004 new reforms of the funding of higher education were announced. The main changes were the reintroduction of maintenance grants in 2004/05 and their substantial increase in 2006/07 and the treble of tuition fees up to a maximum of £3,000 per year in 2006/07. Maintenance loans were only slightly reduced for students whose maintenance grants increased.

[1] The only two countries for which there have been no changes in subsidies for students are Iceland and the Slovak Republic.(OECD (2011))

Economic theory would suggest that increases in tuition fees might affect students' decision to pursue a degree, although this would depend on the relative size of the income and substitution effect associated with increased fees. In particular, if the income effect is larger, higher university costs might reduce enrolment if students do not have the financial resources to cover additional costs. Higher tuition fees, however, might also influence other decisions associated with university choice. Students may, for instance, choose universities charging lower university fees. Alternatively, they may find other cost margins where they can adjust. In particular, alter their decisions associated with university attendance to reduce costs. These could, for instance, be costs associated with living expenses or commuting.

In this paper, we empirically examine how the 2006/07 higher education reform, in which tuition fees trebled in amount, affected English undergraduate students' geographic mobility. Although the reform entailed reforms to tuition fees, loans and grants, which we cannot separately analyze in the data, the major change with respect to the previous policy came from the extent to which tuition fees increase. Moreover, the change in fees applied to all students, while the other reforms were conditional on household income. For simplicity, we refer to the reform as the change in tuition fees. However, in our heterogeneity analysis by income groups we aim to further explore if there were differential effects by income groups, given that students from the bottom of the distribution were eligible for maintenance grants and were also more likely to apply for loans.

Using detailed longitudinal micro-data that follows all students in English state schools from secondary school to university, we look at how the treble of tuition fees up to a maximum of £3,000 per academic year enrolled in university, have influenced students' decisions to leave home to go to university. Specifically, we will consider the geographic distance between a student's home and the university attended, whether the university is located in the same geographical region, as well as the likelihood of enrolling into a university that is located in an affluent area. We also investigate whether there are differences in these decisions associated with attending university across gender and ethnic groups, as well as income distribution.

Understanding geographic mobility within a country is of general interest, since there can be extensive consequences on the economy. Moving away from home to a university situated in another part of the country has been an important feature among young British students. Traditionally, students would select a university based on the rankings for their preferred course of study. If the best matched course was in a university located further away or if other features of said university appealed, it was common practice to move away from home. In this paper, we aim to understand if an increase in fees altered students' choice with respect to the mobility decision.

We use a difference-in-differences approach that compares the change in mobility related to a change in the cost of university among students from different financial backgrounds. We compare students from the very top of the group, who we assume will be largely unaffected by the change in fees, loans and grants, to less well-off students. In addition, to identify the heterogeneity in the response, we compare the reaction throughout the income distribution.

Our results show that the increase in tuition fees did not change the university enrolment rates among English students. There is, however, a sizable effect on geographical mobility. In particular, our findings show that there was a decrease in the distance by 7.0 log points from home travelled by students in the year that the fees changed from a maximum of £1,000 to £3,000 per year. We find, however, that the effect is short-lived. In the second year after the policy reform, there was no significant reduction. This suggests that, for the students in the treatment year, the reform generated some uncertainty, which influenced their university choice and their decision to residing at home or closer to home. Once the reform was better understood, students found ways to adapt to the new funding regime in such a manner that they did not alter their decision to move from home or the distance between home and university.

The treatment heterogeneity analysis shows that students from various income groups were affected by the change, with a more concentrated effect on the students from the bottom distribution when we exclude students who are from London, where there are several universities. Furthermore, our analysis shows that male students reacted to the increased fees more than female students, in terms of reducing the distance they travel. Finally, when we look at differences by ethnicity, we find that, White students reduced the distance to university more than any other non-White students in response to the increase in tuition fees.

We further investigate if students are more likely to pursue a degree in a university located in the same geographical region as their home and to live at home with their parents due to the higher tuition fees. Our findings show that, in the first year of the implementation of the new tuition fee regime, students were more likely to enrol into a university located in the same region as where they reside. Moreover, there is some suggestive evidence that students are more likely to reside at home, although these results were not statistically significant.

Regarding students' mobility across geographic areas by wealth, we show that higher tuition fees triggered a statistically significant decrease in the probability to study in a university located in an area that is economically affluent. The treatment heterogeneity analysis shows that the effects were particularly pronounced among those students at the bottom of the income distribution. Compared with the top 20 percent of the income distribution, the bottom 20 percent, reduced the distance they moved by 4.1 log points.

The rest of the paper is organised as follows. Section 3.2 describes the related literature. Section 3.3 presents the institutional setting and section 3.4 describes the data used. Section 3.5 details the estimation strategy used, while section 3.6 presents the results. Section 3.7 presents the robustness checks and section 3.8 concludes.

3.2 Literature review

There is a growing interest in the impact of changes in university costs on various university outcomes. Our paper relates closely to the literature that studies the effects of increasing tuition fees on students' mobility across state borders. Existing studies have focused on the US, with mixed results. Studies by Tuckman (1970) or Kyung (1996) use state wide data and find that students tend to move more from their home state as the average tuition fees increase in their origin state universities. On the other hand, Dotterweich and Baryla (2005) combine institutional level data for American universities and city level social economic data from 1998. They estimate a differential effect of tuition fees on the enrolment rates of non-resident students by university type. Although there is no statistically significant effect for public universities, higher tuition fees in private tertiary education institutions attract larger inflows of non-resident students. Recent studies have also focused on the case of Germany, where, compared with the US, the level of the tuition fees is very low and most of the universities are public. Dwenger et al. (2012) analyse how the introduction of tuition fees in some German states affected the mobility of medical students. They find that the probability of applying for a university in the home state drops for students who live in a state with tuition fees. Alecke et al. (2013) explore the same German policy but expand the analysis to all students. Their results show a negative effect of tuition fees on enrolment rates only among male students, with first year university students migrating to states that do not charge tuition fees. Related studies exploring the relationship between college costs and educational attainment have also found evidence that lower college costs increase university entry (Deming and Dynarski (2010) review the experimental and quasi-experimental studies in the US; Neill (2009) analyses the Canadian higher education; Hubner (2012) explores the introduction of tuition fees in sixteen German states in 2007).

For the UK, Croxford and Raffe (2014) conduct a descriptive study of the decreasing trend in the flows of English and Northern Irish domiciled students enrolled in Scottish universities after the 2012 British increase in tuition fees. Scotland, unlike the rest of the UK, did not charge fees to home students, however, they applied the same fee structure for students from the rest of the UK. Moreover, unlike the rest of the UK, they offer four rather than three year degree programs, such that the total amount of payable tuition fees is, therefore, much higher. Sa (2014) uses aggregate data to explore variation over time, comparing England and Scotland, to study the effects of changes in tuition on university applications and participation rates. The study shows that an increase in tuition fees decreases the number of university applications, especially for courses with higher earning potential. Enrolment rates also drop, but no evidence of a stronger effect for disadvantaged students is found. Dearden et al. (2011) use data from the British Labour Force Survey between 1992 and 2007 on university participation to analyse the impacts of tuition fees and maintenance grants on university enrolment. They find that a £1,000 increase in fees leads to a drop in participation of 3.9 percentage points, while a £1,000 increase in grants triggers a 2.6 percentage point increase in participation. Our paper contributes to this literature by using detailed panel data to follow students from school to university and study the impact of higher tuition fees on geographical mobility. We can study students' choices to understand the heterogeneity across socio-economic groups, as well as different demographic groups.

Our paper also relates to a literature that studies how geographical distance between students' home and university affects their university and field of study choice. Card (1995) proposes that distance is an important determinant of college participation in the US. Moreover, Frenette (2006) finds that students living very far from university are considerably less likely to enrol than students living ”within commuting distance” in Canada. Spiess and Wrohlich (2010) also document that distance decreases the likelihood to enrol in higher education in Germany. Denzler and Wolter (2011) show that in Switzerland distance between one's home and university affects the choice of field of study and institution, while Gibbons and Vignoles (2012) also find that geographical distance has a significant effect on university choice in England, although it does not affect the decision to enrol.

Finally, through considering the effect of tuition fees on the probability to live with parents during university, our paper is also contributing to the literature on the living arrangement decisions of young people. The studies analysing the factors contributing to the decision of youngsters to delay leaving home have identified income and cultural difference as main drivers. Manacorda and Moretti (2006) find that parents' income is an important determinant of their children's propensity to live at home in Italy: increases in parental income trigger rises in the likelihood of pupils to live with parents. The authors argue that cohabitation is perceived as a normal good by Italian parents. Parisi (2008) also finds that young people from Southern Europe delay leaving home as doing so might increase their chances of being poor. Giuliano (2007) argues that cultural differences have contributed to differential living habits for young people across European countries.

3.3 Institutional setting

In England and Wales different tuition fee regimes for undergraduate degrees have been in place since the late 1990s. This is shown in figure (3.1). The first major change was the introduction of tuition fees in the academic year 1998/99. Students were obliged to pay approximately £1,000 at the beginning of each academic year. However, the actual value of the tuition fee paid by each student depended on their family income. In particular, students were exempt from paying fees if the family income was less than £23,000 per year and were charged a reduced fee if their family income was between £23,001 and £35,000, while those whose families earned more than £35,001 were charged full fees.

The Higher Education Act 2004, effective from 2006/07, changed the regime again through the introduction of variable tuition fees. English universities had discretion over the level of the tuition fees charged, up to a maximum of £3,000 per annum (inflation indexed).[2] Most universities charged the maximum fee permitted. Under the new regime, the level of the tuition fees was not means tested. However, all native students were eligible to apply for tuition fee loans, independent of their economic situation and the value of the loan would cover the entire cost of tuition fee. These loans were payable, in instalments, after graduation and once their income level exceeded £15,000 and the interest rate was very small, close to zero.

The latest change regarding tuition fees was announced in 2010, when the government made public their plans to treble the cap level of tuition fees to £9,000 per year, starting from 2012/13. In our paper, we focus only on the 2006/07 reform.

[2] Devolution meant that Scotland, Northern Ireland and Wales pursued different policies.

In conjunction with increased tuition fees and the introduction of maintenance loans, amendments were made to means-tested grants. Specifically, the maintenance grants which were halved in 1998 and then abolished in 1999 were reintroduced in 2004/05. Initially, the maximum value was of up to £1,040 per year for the most deprived students and it was further increased to a maximum of £2,700 in 2006/07. In parallel, available maintenance loans increased in 1998: for those previously eligible for maintenance grants the increase was similar to the reduction in the grant and for those liable for the new fees the increase was similar to the increase in the tuition fees. Between 2004/05-2006/07 maintenance loans remained stable, though they were reduced slightly for students who benefited from maintenance grants. It is worth noting that the value of these maintenance loans depended on the location of the university and whether the student lived at home with their parents. In particular, those studying in a London university benefited from larger sums and those living at home were eligible for lower maintenance loans.

Table (3.1) summarises all the fees and the financial support available to students based on their family income level both before and after the 2006/07 change in tuition fees, which is the policy we investigate in the paper. We present figures for the first year for which we have available data (2004/05) and the first year with the new policy in place. Although the tuition fees, tuition fee loans and maintenance grants are the same for all students independent of their living arrangements and university location, the last two columns which show the maximum available maintenance loan a student can apply for, refer to students not pursuing a degree in London and not living at home with their parents.

3.4 Data and descriptive statistics

In this section, we start by describing the main data sources used in the analysis. We then proceed by presenting some summary statistics and describing the main outcome variables.

3.4.1 Data

In the analysis we use individual-level data linking information from the National Pupil Database (NPD) and Higher Education Statistical Agency (HESA).[3] The NPD is provided by the English Department for Education and comprises an administrative data set of all pupils enrolled in state schools in England.[4] We focus on students enrolled in secondary education and use mainly information contained in the Pupil Level Annual School Census (PLASC), which is one of the many data sets included in the NPD. In particular, we use detailed information on the geographical residence of pupils (we have information at lower layer super output area level, totalling around 32,400 areas), variables related to demographic characteristics (for instance, gender and ethnic origins), as well as students' grades obtained at a national level examination, called General Certificate of Secondary Education (GCSE).[5] All variables are measured when pupils finish compulsory education (i.e. when individuals are aged 16). This information is linked to a second administrative data set, HESA, which contains information on students pursuing an undergraduate degree in English universities.

This linked data set allows us to follow all students in English states schools from secondary school to university. Our analysis is based on information on first year undergraduate English students entering university between 2004/05 - the first year of data available - and 2007/08. We consider data only until 2007/08 in order to abstract from the effects of the financial crisis on university participation. In our empirical analysis, the first two years of data (2004/05-2005/06) represent the period before the change in fees, while the years 2006/07-2007/08 represent the period after the new fee regime that capped tuition fees at £3,000 was in effect from 2006/07. Moreover, we restrict our analysis to students who enrol in the year in which they are eligible to attend university.

3.4.2 Summary statistics

The empirical analysis compares the effect of a change in tuition fees on students from different social economic status. We assume that students from very wealthy families will be largely unaffected by the change in tuition fees or maintenance grants and loans, while other students might respond. The intuition being that students from families at

[3] See chapter 2 for further details on the linked NPD and HESA data.

[4] In England, 93% of pupils are enrolled in state schools.

[5] The lower layer super output area covers areas with minimum 1,000 (400) and maximum 3,000 (1,200) individuals (households). There are in total 32,482 lower layer super output area in England in the period we consider.

the top of the income distribution are unlikely to change their decision to go to university, or any related decision (for instance, preferred university or degree choice). In particular, we construct the treatment and control group using the students' Income Domain Affecting Children Index (IDACI). This indicator is a continuous variable between 0 and 1 that measures the percentage of children aged 0-15 years old living in income-deprived families in lower layer super output area.[6] For each cohort of pupils finishing their compulsory school we group pupils based on the percentiles of the IDACI score. Then, for those that join tertiary education we define the control group as those in the first percentile of their cohort and the treated group the remaining students. We later categorise pupils into income groups to understand the heterogeneity in response throughout the distribution.

As table (3.2) shows, in the control group, on average, students live in areas where there are almost no pupils aged 0-15 that are coming from deprived families. Thus, it seems that our control group includes the richest children in our data set. Students in the treatment group are indeed from poorer backgrounds, as they live in areas where around 16-17% of households are income deprived. Moreover, there is a large variation in the share of income deprived households in the area where pupils live within the treatment group, with an average ranging from around 4%, for the top of the distribution, to 50%, for the bottom of the distribution.

Given that one of the main components of the 2006/07 was the treble of tuition fees, we further analyse how the financing sources for tuition fees differ by treatment group before and after the treble of the tuition fees. Table (3.3) shows that before the 2006/07 increase in tuition fees, 70% of the students in the control group were paying the tuition fees using their own or their family's financial resources, while among the treated students only 50% were paying their fees without using any additional financial help. Furthermore, among the poorer students from the treated group only 30% paid their fees on their own. Once the fees reached a maximum of £3,000 both students from the control and the treated group made use of the tuition fees loans, as, on average, 70% of each group got the loan offered by the Student Loan Company a non-profit making, government-owned, organisation.[7]

[6] It should be noted that a household is considered to be income deprived if the household income (before housing costs and without housing benefits) is below 60% of the national median income and if they are receiving any form of income support or benefits. Source: Association of Public Health Observatories, 2012 Deprivation scores. Website: `http://www.makingthelink.net/data-source/deprivation-scores`

[7] This includes awards assessed by English or Welsh Student Award Agency and paid in full by the Student Loan Company (SLC), whether it is SLC full funded fees through a grant or through a fee loan or through a mixture of SLC grant and SLC fee loan.

Table (3.4) sheds light on the characteristics of our control and treated group before the 2006/07 reform. In panel A we look at demographic characteristics and the academic performance of students at the exam taken at the end of compulsory education, the GCSE. It seems that female students are slightly more present in the treated group compared to the control. Moreover, around 95% of students in the control group are White, while the ethnic minorities represent around 17% in the treated group, with those of Asian origin being the larger minority. As described in chapter 2, all students take the GCSE in English and Mathematics. In our sample, students in the control group have higher grades in both English and Mathematics in the control group, compared to the treated group.

Panel B presents the outcome variables we are considering in the analysis. Our main outcome variable is the geographical distance between a student's home address at age 16 and the university attended. In order to calculate this distance we use the coordinates of the centroid of the lower layer super output area, which is the most disaggregated geographical location we have access to, and the geographical coordinates of the university's postcode. On average, students in the control group attend universities that are farther away from home compared to the ones from the treated group, with the former travelling around 80 kilometres and the latter 46 kilometres on average. Furthermore, students in the treated group are also more likely to pursue a degree in the same geographical region as their home and also more prone to live at home with their parents relative to the control group.[8]

We also consider the wealth of the area in which the university attended is located. We use average house prices, measured in each lower layer super output area in the third quarter of each year, which is provided by the Office of National Statistics.[9] We divide the house prices by quartiles and define the outcome variable as the probability to live in an area with house prices above the median. The probability that students from the control group study in a university located in a rich area is slightly larger (55%) compared to the probability for the treated group (52%). This is not surprising given that students in our control group are more prone to be living in areas with relative high house prices already when they are 16.

[8]We use the ONS definition of English geographical regions: East of England; East Midlands; London; North East; North West; South East; South West; West Midlands.

[9]The Office of National Statistics (ONS) provides median house prices by middle layer super output area for each quarter. We link this data to the lower layer super output area using the mapping data set provided by the ONS. In order to keep prices constant, we also use the 2004 retail price index.

3.5 Empirical strategy

Ideally, to estimate a causal effect of the increase in tuition fees on students' mobility, we would observe the same individual both before and after the reform, and analyse how the higher tuition fees impacted their mobility. Since this is not possible, we use a difference-in-differences approach by comparing the change in the mobility of students affected by policy change ("treated" group) to the change in the mobility of a ("control") group that is unlikely to have been affected by the policy change, but who are also attending university in the same academic year.

We define the control group as the richest pupils as measured when they are 16 using the top percentile according to the IDACI score. As we showed in section 3.4 this corresponds to around 1% of the pupils aged 0-15 in the lower layer super output area coming from deprived families. We further define the treatment group as the rest of the pupils. The main equation estimated in the analysis is the following:

$$y_{it} = \alpha_1 + \alpha_2 T_i + \alpha_3 T_i A_t + X_i \theta + d_t + \varepsilon_{it} \tag{3.1}$$

where y_{it} is the outcome variable (for instance, the logarithm of the geographical distance between home and the university attended or the probability to study in a university located in a rich area). T_i is a treatment dummy equal to 1 if individual i belongs to the treatment group and 0 otherwise and it controls for differences in the outcome variable between students in the control and the treated group. A_t is a dummy equal to 1 if individual i is enrolled as a first year student in year t after the change in the reform (i.e. either in 2006/07 or 2007/08) and 0 if they are enrolled in one of the two years before the change in the tuition fee regime (i.e. either in 2004/05 or 2005/06); X_i represents individual characteristics (gender, ethnicity, grades at examination taken at the end of compulsory education); d_t stands for time fixed effects and controls for changes over time in the outcome variable. ε_{it} is an idiosyncratic error term.

The causal effect of change in tuition fees on the geographical distance between a student's home and the university attended is captured by α_3 and can be interpreted as the change (in log points) in the distance travelled induced by the reforms. Our identification strategy is justified in part by the results presented in section (3.6.1), where we show that enrolment rates are unaffected by the 2006/07 change in tuition fees. If enrolment had been discouraged by the reform it might suggest changes in the composition of the pool of students attending university. Furthermore, enrolment rates remaining unchanged allow us to continue to observe the university choice made by students, which is key in identifying our outcome variable (i.e., the geographical distance).

In our analysis, we consider as well how the effect changes in time after the introduction of the new tuition fees. Given that we use data only until 2007/08, we can analyse the effect only in the year with the change in the data and the following year:

$$y_{it} = \beta_1 + \beta_2 T_i + \beta_3 T_i A_{06} + \beta_4 T_i A_{07} + X_i \theta + d_t + \varepsilon_{it} \tag{3.2}$$

with A_{06} and A_{07} being categorical variables equal to 1 if student i is starting university in 2006/07 and 2007/08, respectively.

Finally, given that the treatment group includes pupils from a wide range of social economic status, we investigate if there are heterogeneous effects by treatment subgroups:

$$y_{it} = \gamma_1 + \sum_j \gamma_j T_{ij} + \sum_j \omega_j T_{ij} A_t + X_i \theta + d_t + \varepsilon_{it} \tag{3.3}$$

where j takes values $2-20$, $21-40$, $41-60$, $61-80$ and $81-100$; T_{ij} are dummy variables equal to 1 if individual i is in the 2-20th, 21-40th, 41-60th, 61-80th and 81-100th income distribution, respectively, and 0 otherwise; ω_j captures the heterogeneous effects for each treatment group j.

The identifying assumption of the difference-in-differences estimation is that the contemporaneous trend in the outcome variable would have been the same for the rich and the poor children (our control and treatment group) in the absence of the change in tuition fees. The new levels of the tuition fees might trigger a deviation from this trend and the difference-in-differences estimator would measure this deviation. Graph (3.2), which plots the average distance between home and university by treatment and across time, shows that for the treatment group there is a sharp decrease in the average distance travelled in 2006/07, the first year of the new tuition fee regime. Regarding the control group, the average distance between university and home is quite stable across time. This suggests that our common trends assumption holds.

We further test for different time trends by conducting a placebo test in which we define the post-period as the academic year 2005/06, the last year before the change in tuition fees, and the pre-period as the first academic year for which we have data, 2004/05. We report our results in table (3.5). Columns (1) and (3) present the overall effect and columns (2) and (4) investigate whether there are any heterogeneous effects within the treated group by estimating equation (3.3). Our estimates show that there is no statistically significant difference in the distance to university between the treatment and control group, both overall and when we consider the heterogeneity within the treatment group. The results do not change if we consider or not those students who live in London. Thus, we can conclude that our falsification test brings further evidence that our assumption of common trends holds.

3.6 Results

In this section we present our main results. We begin by analysing how the higher tuition fees affected enrolment, as one potential effect is a change in the number of students enrolled after the higher tuition fees were introduced in 2006/07. This might also suggest a change in the composition of students attending university before and after the reform. We then proceed by analysing the effect on the distance between home and university, which is our main measure of mobility. We also investigate if there are heterogeneous effects along the income distribution, as well as by demographic characteristics. We further study changes in mobility decisions by studying the effects of the larger tuition fees on the likelihood to pursue a degree in the same geographical region as one's home at age 16, the likelihood to live at home with parents while studying at university and on the probability to study in a university located in a rich area.

3.6.1 Enrolment

We estimate equation (3.1) to measure the effect of the change in tuition fees on enrolment. The outcome variable is defined as a categorical variable, which takes value 1 if individual i is enrolled into any university and 0 otherwise. Column (1) of table (3.6) reports the estimate of α_3. It shows that the higher tuition fees have no statistically significant effect on enrolment rates. Moreover, the point estimate is small. When we control for the demographic characteristics, the estimate of α_3 is negative, although not statistically significant. We find that female students are more likely to pursue an undergraduate degree compared to male. Moreover, ethnic minorities are also more likely to enrol into university, as are high-performing students. Overall, the analysis suggests that the changes in tuition fees did not alter the composition of university entrants, such that, over the time period in our study, the cohorts are comparable.

In addition, we estimate the year-on-year response to the change in in fees in 2006. Columns (3) and (4) report the estimates of β_3 and β_4 from equation (3.2). The findings show that there is no differential effect on enrolment among English secondary school graduates over time.

3.6.2 Distance - overall effect

We now proceed to estimate the impact of the higher tuition fees on the logarithm of the geographical distance between a student's home address and the university they attend.

Table (3.7) presents the results. The first six columns refer to all students enrolled, while in the last six columns the sample is restricted only to those students who were not living in London. Columns (1) and (8) estimate equation (3.1) without controlling for student characteristics. The analysis shows that overall there is no statistically significant effect of the change in tuition fees on distance. This is the case for both the entire sample and when we restrict the sample to exclude London-based students. When we add the student characteristics, the magnitude of the effect drops slightly. It seems that female and ethnic minorities are less likely to travel long distance between home and the university attended compared to male and Whites, respectively.

Focusing on the year-on-year analysis, however, we do find a strong short-run effect on distance. Columns (3), (4), (9) and (10), which estimate equation (3.2), show the effect of the fees is strong and significant in the year when the changes were first introduced. Our findings suggest that the higher tuition fees and maintenance grants triggered a decrease in the distance travelled of 7.0 log points and of 6.8 log points among those not living in London. This suggests that the change in regime prompted a reaction by the first cohort of students affected by the reform, however, after one year, students adapted to the new regime, such that, despite the same fee regime, they followed similar patterns in university choice to those in the pre-reform period.

In order to understand if there are heterogeneous effects across socio-economic groups within the treatment group, we estimate equation (3) separately by income group. The results are presented in columns (5), (6), (11) and (12). For the post reform period, we focus only on the first year after the introduction of the higher fees. The treatment heterogeneity analysis shows that students from all income groups are affected by the change. The largest effect, however, is among students from the bottom distribution who do not live in London. In particular, we find that for students in the bottom 20th percentile of the distribution and who do not reside in London, the higher tuition fees decreased the distance travelled by 8.6 log points.

3.6.3 Distance - heterogeneity by student demographics

We further analyse how the effect of the increased tuition fees on distance between home and university differs by demographic characteristics. We begin by looking at gender and then ethnicity. We restrict our attention to the period immediately after the change in fees, as this is where we see the strongest reaction. In order to estimate these heterogeneous effects we use equation (3.1) and we restrict the sample to each demographic group of English students.

The first six columns of table (3.8) present results for all native students, while columns (7) - (12) consider only students not living in London. In columns (1), (2), (7) and (8) we split the sample by gender. Our estimates show that, while there is no effect for female students, male students choose a university closer to home than before the change in fees. In particular, the reform reduces the distance travelled by male students by 10 log points in the overall sample for English undergraduates not living in London. In other words we find gender differences in the reaction to the higher tuition fees, with men reducing the financial burden introduced by the higher tuition fees by enrolling into universities closer to home. Our findings are not supporting the existing empirical findings that men are less risk averse than women (see the surveys presented in Eckel and Grossman (2008) and Croson and Gneezy (2009).

Given the increasing ethnic diversity within England, in columns (3) - (6) and (9) - (12) we consider the heterogeneity of the effect across different ethnic groups. Our findings suggest that the increase in fees in 2006/07 reduced the distance travelled by White students by 7.3 log points in the overall sample and for students living outside London. Other ethnic groups (Black, Asian and others), however, did not react to the change in the fee regime by selecting a university closer to home. Our results point to cultural differences although we would have expected that ethnic minorities, who are more likely to be sensitive to problems of geographical mobility due to very close family ties (especially Asians), to be more responsive to the policy change than Whites.

3.6.4 Regional mobility

In this subsection we analyse the effect of a change in fees on the likelihood to select a university that is in the same region as their home and also on the likelihood to reside at home when at university.

Table (3.9) shows that higher tuition fees increase the probability to pursue a degree in a university located in the same region as a student's home. Our findings show that the effects are statistically significant in the short run, with the new regime increasing the likelihood to study in the one's region by 2.4 log points for the entire sample and 2.6 log points for students not living in London. Moreover, columns (3) and (6) indicate that the effect is mainly triggered by students in the top and bottom distribution, with larger magnitudes for the latter group.

In table (3.10) we consider the impact of the change in tuition fees on the likelihood to live with parents, while enrolled as a first year undergraduate student. The positive estimates presented in columns (1) - (6) suggest an increase in the probability to live at home with parents, although none of the estimates is statistically significant either overall, by year or by heterogeneity of the treatment, whether we consider students who live in London or not.

3.6.5 Wealth of area

We now proceed by analysing if students not only go to universities that are closer to home, but also choose universities located in areas that are affluent. This is an interesting aspect to consider since we would expect affluent areas to have stronger local labour markets, which might benefit students when whey graduate from university. However, these areas are also likely to be more expensive to live in the short-run. It, therefore, triggers a trade-off between short-term costs and (potentially) long-run gains. We proxy for local labour market prosperity using the area's house prices. We define the outcome as the probability to attend a university located in an area with house prices above the median.

Columns (1) and (4) of table (3.11) show that, on average, in the post-reform period, there is no overall effect. However, as the results reported in column (2) indicate, there is a short run effect, with students being less likely to enrol into universities that are located in rich areas in the first year of the new policy. The effect disappears in the following year as well as when we restrict the analysis to students not residing in London, although the negative signs of the estimates suggest that students are less likely to enrol into universities located in affluent areas. The results suggest that students, overall, are staying closer to home, and, when they move, they are more likely to move to areas that are, on average, less affluent.

Columns (3) and (6), indicate, however, that there are stark differences in the areas where students locate, depending on their economic backgrounds. We find that, students from wealthier backgrounds (i.e. those in the 2-40% top percentiles) do not alter the area where they locate when the funding of higher education regime change, while the poorest group of students are less likely to pursue a degree in a rich area. In particular, we find that higher tuition fees decrease the probability of lower income students to enrol into a university located into an affluent area by 4.1 log points. Overall, we interpret these results as suggestive that the mobility of the poorest students is the most affected by the change in fees. Not only that they are less likely to attend a university that is located further away from home, but they are also less likely to choose a university located in an area that has a stronger local labour market.

3.7 Robustness checks

In this section, we test the robustness of our results by considering a series of alternative specifications for our main variable of interest: the geographical distance between home and university. We consider three different measures for the control group. In panel A of table (3.12) we define the control group as the top 1.5% percentile of the IDACI score distribution. Comparing to our main results presented in table (3.7) it seems that including students from poorer areas in our control group decreases the magnitude of the estimates, both on average and in the year on year analysis. Moreover, it seems that those students not living in London trigger the main effect.

In panels B and C we include even poorer students in the control group, which we define as the top 2% percentile of the IDACI score distribution and the top 2.5% percentile of the IDACI score distribution, respectively. The magnitude of the effects reported in columns (1) and (2) decreases even more as we include students from the lower IDACI score distribution in the control group.

3.8 Conclusion and future work

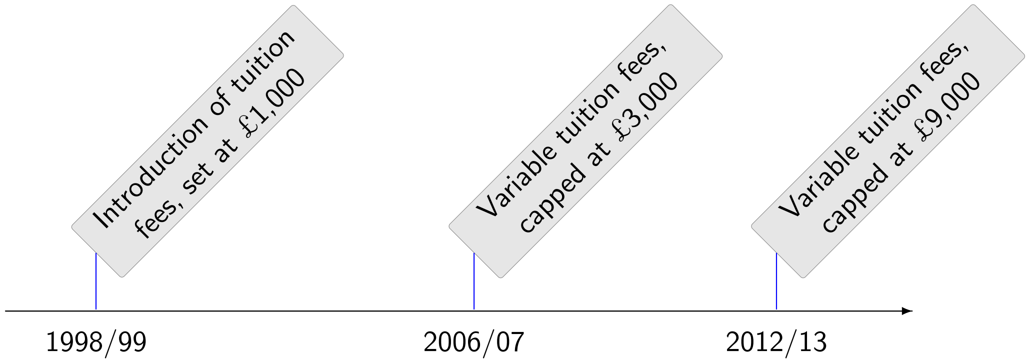
The intended (and unintended) consequences of the introduction of, and increase in, tuition fees in England and Wales, are still relatively unclear. In this paper, we use a difference-in-differences estimation to estimate the effect of the increase in tuition fees charged by English university in 2006/07 on English students' mobility. Our findings suggest that increases in tuition fees did not impact the extensive margin (i.e., the university enrolment rates) among English students, but it influenced factors relating to university choice. We focus on students' geographical mobility and find that the distance travelled by students decreased by 7.0 log points in the first year when the reform was implemented. In other words, given that students pursue an undergraduate degree in a university located on average approximately 50 kilometers away from home, the increase in tuition fees and maintenance grants decreased the distance travelled by approximately 7 kilometers on average. Moreover, we find that students are more likely to pursue a degree in a university located in the same geographical region as their home. These effects, however, were short-lived, and the following year, there were no significant differences.

The treatment heterogeneity analysis in the short-run reveals that students from various income groups were affected by the change, although the larger effect is found for students from the bottom distribution when we exclude students who are from London. Furthermore, our analysis by demographic groups suggests that male students, as well as White students are travelling smaller distances due to the increase in tuition fees. We further test the likelihood of students to locate in an affluent area and find that this is reduced, especially for those at the bottom of the wealth distribution.

We plan to expand on this paper in a number of directions. First, given the detailed data we have for students and schools at the neighbourhood level, we plan to improve our estimation strategy by matching students and re-estimating our analysis, before and after the change tuition fees. Second, we hope to allow for an additional dimension of control using linked school and university data from Scotland, where there was no change in the fee regime over the same period. Third, we plan to measure the impact of the introduction of fees on other important dimensions of university choice. For instance, we aim to look at the ranking of university, as well as the course choice. Forth, by acquiring additional data, we intend to investigate the more drastic change in fees that took place in 2012-13, whereby universities were entitled to charge a maximum of £9,000. Finally, our future plan is to study the long run effects of higher tuition fees on labour market outcomes using the Destinations of Leavers from Higher Education, a survey conducted six months after graduation.

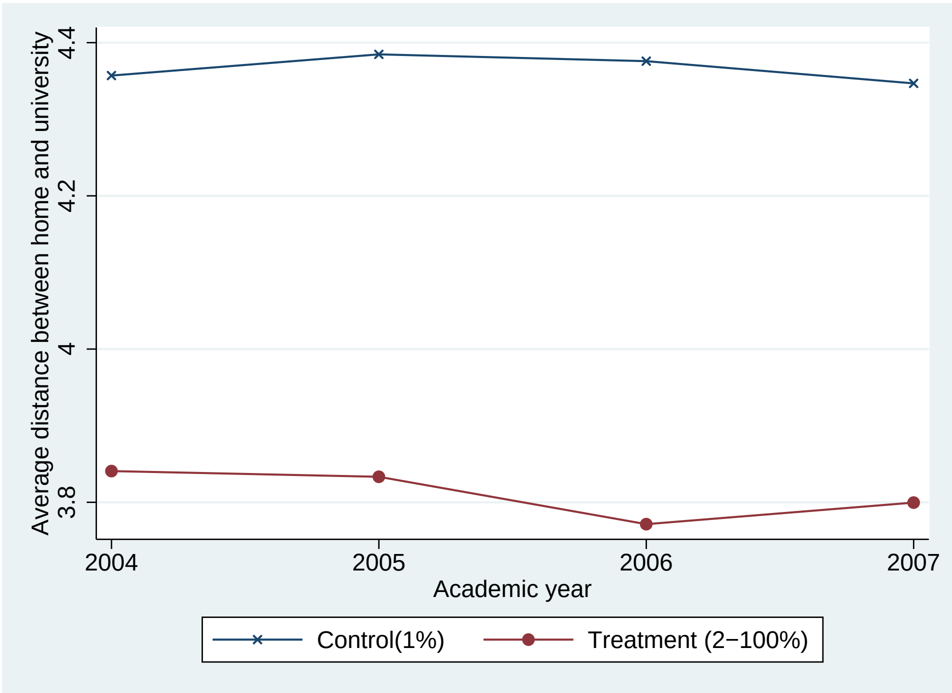
Tables and figures

FIGURE 3.1: Time line of tuition fees in England



Notes: The time line presents changes in tuition fees charged by English universities for undergraduate degrees between 1998/99-2012/13.

FIGURE 3.2: Parallel trends: distance between home and university



Notes: The graph depicts the average logarithm of distance between home and universities for students in the control and in the treatment group, between 2004/05-2007/08. Source: Authors' calculations using linked NPD-HESA data.

TABLE 3.1: Available tuition and maintenance financial support by income brackets

Parental income (£)	Tuition fees (£)		Tuition fees loans (£)		Maintenance grants (£)		Maximum maintenance loans (£)	
	2004/05	2006/07	2004/05	2006/07	2004/05	2006/07	2004/05	2006/07
$\leq$ 10,000	0	3,000	0	3,000	1,040	2,700	4,260	3,205
20,000	0	3,000	0	3,000	248	2,283	4,260	3,205
30,000	980	3,000	0	3,000	0	832	4,260	4,037
40,000	1,196	3,000	0	3,000	0	0	3,262	4,183
50,000	1,196	3,000	0	3,000	0	0	3,199	3,305

Notes: The figure for tuition fee loans, maintenance grants and maintenance loans refer to students who do not study in London and do not live at home with their parents. Source: Student Loan Company

TABLE 3.2: Average IDACI score by period and treatment

Control	Treated	Treated 2-20%	Treated 21-40 %	Treated 41-60%	Treated 61-80 %	Treated 81-100%
Panel A: Period 2004/05-2005/06						
0.010	0.158	0.036	0.082	0.157	0.282	0.498
(0.002)	(0.151)	(0.012)	(0.016)	(0.028)	(0.046)	(0.105)
Panel B: Period 2006/07-2007/08						
0.010	0.169	0.036	0.083	0.157	0.282	0.501
(0.002)	(0.158)	(0.012)	(0.016)	(0.028)	(0.046)	(0.108)

Notes: Panel A refers to the two years before the change in tuition fee, while panel B refers to the period after the change in tuition in fees. The first row in each panel reports the average IDACI score. Standard deviations reported in parentheses. Source: Authors' calculations using the linked NPD-HESA data

TABLE 3.3: Major source of tuition fees funding by period and treatment

	Control	Treated	Treated 2-20%	Treated 21-40 %	Treated 41-60%	Treated 61-80 %	Treated 81-100%
Panel A: Period 2004/05-2005/06							
No award or financial backing	69.27%	49.31%	60.19%	53.14%	47.13%	38.09%	29.46%
Award assessed by Student Award Agency	21.26%	35.19%	25.93%	30.40%	35.82%	45.36%	56.98%
Other award sources	9.47%	15.40%	13.88%	16.46%	17.05%	16.55%	13.56%
Panel B: Period 2006/07-2006/07							
No award or financial backing	23.13%	21.4%	20.92%	20.18%	20.46%	21.6%	25.88%
Award assessed by Student Award Agency	69.76%	70.19%	70.89%	71.19%	70.70%	69.99%	66.26%
Other award sources	7.10%	8.42%	8.19%	8.63%	8.84%	8.41%	7.86%

Notes: Panel A refers to the two years before the change in tuition fee, while panel B refers to the period after the change in tuition in fees. Each row shows the percentage of students for whom the major source of financial resources is: their own financial resources; the Student Award Agency which includes awards assessed by English or Welsh Student Award Agency and paid in full by Student Award Agency or by the Student Loan Company (SLC) when the SLC fund full fees through a grant or when the SLC fund full fees through a fee loan or when the SLC fund full fees through a mixture of SLC grant and SLC fee loan; other financial support. Source: Aauthors' calculations using the linked NPD-HESA data

TABLE 3.4: Descriptive statistics by treatment, before the change in tuition fees

	Control	Treated
Panel A: Demographics and academic performance		
Female	0.546	0.572
	(0.498)	(0.495)
White	0.946	0.823
	(0.227)	(0.381)
Asian	0.033	0.114
	(0.179)	(0.318)
Black	0.004	0.034
	(0.061)	(0.180)
Other ethnic group	0.017	0.029
	(0.131)	(0.169)
Standardised GCSE English grade	0.966	0.822
	(0.453)	(0.492)
Standardised GCSE Mathematics grade	1.049	0.853
	(0.512)	(0.552)
Panel B:Outcome variables		
ln Distance	4.377	3.835
	(0.969)	(1.272)
Probability to study in the same region	0.318	0.484
	(0.466)	(0.500)
Probability to live with parents	0.103	0.250
	(0.304)	(0.484)
Probability to study in a rich area	0.552	0.515
	(0.497)	(0.500)
Observations	8,349	398,274

Notes: Mean reported. Standard deviations reported in brackets. Source: author's own calculations using the linked NPD-HESA data

TABLE 3.5: Placebo test

	All students		Only students not living in London	
	Overall	Treatment Heterogeneity	Overall	Treatment Heterogeneity
	(1)	(2)	(3)	(4)
After * Treatment	-0.027		-0.018	
	(0.030)		(0.031)	
After 2006 * Treatment(2-20)		-0.007		-0.006
		(0.031)		(0.031)
After 2006 * Treatment(21-40)		-0.029		-0.018
		(0.031)		(0.032)
After 2006 * Treatment(41-60)		-0.022		-0.012
		(0.032)		(0.033)
After 2006 * Treatment(61-80)		-0.044		-0.039
		(0.033)		(0.035)
After 2006 * Treatment(81-100)		-0.049		-0.053
		(0.036)		(0.040)
Female	-0.099***	-0.085***	-0.095***	-0.082***
	(0.006)	(0.006)	(0.006)	(0.006)
Asian	-0.834***	-0.626***	-0.689***	-0.506***
	(0.010)	(0.010)	(0.013)	(0.013)
Black	-0.441***	-0.097***	-0.231***	-0.013
	(0.017)	(0.018)	(0.033)	(0.033)
Other ethnicity	-0.464***	-0.294***	-0.179***	-0.090***
	(0.019)	(0.018)	(0.025)	(0.024)
Standardised GCSE English grade	0.310***	0.279***	0.316***	0.287***
	(0.006)	(0.006)	(0.007)	(0.007)
Standardised GCSE Mathematics grade	0.331***	0.284***	0.334***	0.291***
	(0.006)	(0.006)	(0.006)	(0.006)
Observations	194,751	194,751	166,158	166,158
Time FE	X	X	X	X

Notes: In regressions in columns (1)-(2) the sample includes all students. In regressions (3) -(4) the sample is restricted only to students who are not residing in London. The outcome variable is the logarithm of the geographical distance between home and the university attended. All regressions include the treatment dummy as well and cover period 2004/05-2005/06. The after period is defined as the academic year 2005/06 and the before period is defined as the academic year 2004/05. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE 3.6: Probability to enrol into university

	Overall	Overall	Year on Year	Year on Year
	(1)	(2)	(3)	(4)
After * Treatment	0.002	-0.000		
	(0.007)	(0.006)		
After 2006 * Treatment			0.003	0.003
			(0.009)	(0.007)
After 2007 * Treatment			0.001	-0.003
			(0.009)	(0.007)
Female		0.029***		0.029***
		(0.001)		(0.001)
Asian		0.222***		0.222***
		(0.001)		(0.001)
Black		0.153***		0.153***
		(0.002)		(0.002)
Other ethnicity		0.098***		0.098***
		(0.002)		(0.002)
Standardised GCSE Mathematics grade		0.137***		0.137***
		(0.000)		(0.000)
Standardised GCSE English grade		0.112***		0.112***
		(0.000)		(0.000)
Time FE	X	X	X	X

Notes: The outcome variable is a dummy equal to 1 if enrolled into any university with or without a gap year between 2004/05-2011/12, and zero otherwise. Both regressions control for the treatment dummy as well. Robust standard errors in parentheses.*** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 3.7: Effect on distance between home and university

	All students						Only students not living in London					
	Overall		Year on Year		Treatment Heterogeneity		Overall		Year on Year		Treatment Heterogeneity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
After * Treatment	-0.031	-0.027					-0.029	-0.028				
	(0.022)	(0.021)					(0.022)	(0.021)				
After 2006 * Treatment			-0.073***	-0.070***					-0.066**	-0.068***		
			(0.026)	(0.026)					(0.026)	(0.026)		
After 2007 * Treatment			0.010	0.016					0.008	0.013		
			(0.027)	(0.026)					(0.027)	(0.026)		
After 2006 * Treatment(2-20)					-0.070***	-0.065**					-0.064**	-0.062**
					(0.027)	(0.026)					(0.027)	(0.026)
After 2006 * Treatment(21-40)					-0.068**	-0.070***					-0.062**	-0.068**
					(0.027)	(0.027)					(0.028)	(0.027)
After 2006 * Treatment(41-60)					-0.061**	-0.066**					-0.051*	-0.058**
					(0.028)	(0.027)					(0.029)	(0.028)
After 2006 * Treatment(61-80)					-0.038	-0.047*					-0.038	-0.049*
					(0.029)	(0.028)					(0.030)	(0.029)
After 2006 * Treatment(81-100)					-0.052*	-0.065**					-0.071**	-0.086***
					(0.031)	(0.030)					(0.034)	(0.033)
Female		-0.099***		-0.100***		-0.085***		-0.097***		-0.097***		-0.081***
		(0.004)		(0.004)		(0.004)		(0.004)		(0.004)		(0.005)
Asian		-0.822***		-0.822***		-0.631***		-0.674***		-0.674***		-0.508***
		(0.007)		(0.007)		(0.008)		(0.009)		(0.009)		(0.010)
Black		-0.418***		-0.418***		-0.098***		-0.244***		-0.244***		-0.010
		(0.011)		(0.011)		(0.014)		(0.023)		(0.023)		(0.027)
Other ethnicity		-0.405***		-0.405***		-0.276***		-0.152***		-0.152***		-0.081***
		(0.012)		(0.012)		(0.014)		(0.016)		(0.016)		(0.019)
Standardised GCSE English grade		0.339***		0.339***		0.294***		0.343***		0.343***		0.300***
		(0.005)		(0.005)		(0.005)		(0.005)		(0.005)		(0.006)
Standardised GCSE Mathematics grade		0.343***		0.343***		0.292***		0.351***		0.351***		0.302***
		(0.004)		(0.004)		(0.005)		(0.004)		(0.004)		(0.005)
Observations	406,623	406,623	406,623	406,623	299,755	299,755	345,110	345,110	345,110	345,110	254,710	254,710
Time FE	X	X	X	X	X	X	X	X	X	X	X	X

Notes: In regressions in columns (1)-(6) the sample includes all students. In regressions (7) -(12) the sample is restricted only to students who are not residing in London. The outcome variable is the logarithm of the geographical distance between home and the university attended. Regressions (1), (2), (4) and (5) include the treatment dummy as well and cover period 2004/05-2007/08. Regressions (3) and (6) include the individual treatment dummies and cover periods 2004/05-2006/07. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE 3.8: Effect on distance between home and university - heterogeneity by student demographics

	All students						Only students not living in London					
	Female	Male	White	Asian	Black	Other	Female	Male	White	Asian	Black	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
After 2006 * Treatment	-0.045	-0.100***	-0.073***	0.076	-0.439	-0.023	-0.039	-0.104***	-0.073***	0.077	-0.404	0.082
	(0.035)	(0.037)	(0.026)	(0.153)	(0.296)	(0.184)	(0.036)	(0.038)	(0.027)	(0.153)	(0.367)	(0.187)
Female			-0.089***	-0.166***	0.002	-0.134***			-0.085***	-0.197***	0.053	-0.107***
			(0.005)	(0.015)	(0.028)	(0.031)			(0.005)	(0.020)	(0.057)	(0.040)
Asian	-0.868***	-0.787***					-0.740***	-0.620***				
	(0.010)	(0.012)					(0.014)	(0.015)				
Black	-0.400***	-0.513***					-0.177***	-0.334***				
	(0.017)	(0.023)					(0.034)	(0.045)				
Other ethnicity	-0.442***	-0.434***					-0.167***	-0.168***				
	(0.020)	(0.023)					(0.026)	(0.030)				
Standardised GCSE English grade	0.356***	0.295***	0.333***	0.252***	0.293***	0.462***	0.365***	0.294***	0.331***	0.299***	0.377***	0.446***
	(0.007)	(0.008)	(0.006)	(0.017)	(0.031)	(0.034)	(0.008)	(0.008)	(0.006)	(0.023)	(0.062)	(0.048)
Standardised GCSE Mathematics grade	0.338***	0.338***	0.347***	0.364***	0.201***	0.260***	0.341***	0.348***	0.334***	0.461***	0.298***	0.319***
	(0.006)	(0.007)	(0.005)	(0.015)	(0.025)	(0.030)	(0.007)	(0.008)	(0.005)	(0.020)	(0.051)	(0.041)
Observations	171,398	128,357	248,475	33,682	9,401	8,197	145,316	109,394	227,378	20,080	2,724	4,528
Time FE	X	X	X	X	X	X	X	X	X	X	X	X

Notes: In regressions in columns (1)-(6) the sample includes all students. In regressions (7) -(12) the sample is restricted only to students who are not residing in London. The outcome variable is the logarithm of the geographical distance between home and the university attended. Each column refers to a different group of natives, as specified in the title. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE 3.9: Effect on the probability to study in the same region

	All students			Only students not living in London		
	Overall	Year on Year	Treatment Heterogeneity	Overall	Year on Year	Treatment Heterogeneity
	(1)	(2)	(3)	(4)	(5)	(6)
After * Treatment	0.006			0.007		
	(0.010)			(0.010)		
After 2006 * Treatment		0.024**			0.026**	
		(0.012)			(0.013)	
After 2007 * Treatment		-0.012			-0.011	
		(0.012)			(0.013)	
After 2006 * Treatment(2-20)			0.022*			0.023*
			(0.013)			(0.013)
After 2006 * Treatment(21-40)			0.026**			0.027**
			(0.013)			(0.013)
After 2006 * Treatment(41-60)			0.019			0.019
			(0.013)			(0.013)
After 2006 * Treatment(61-80)			0.022*			0.025*
			(0.013)			(0.014)
After 2006 * Treatment(81-100)			0.025*			0.036**
			(0.013)			(0.014)
Female	0.038***	0.038***	0.034***	0.038***	0.038***	0.033***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Asian	0.128***	0.128***	0.072***	0.045***	0.045***	-0.006*
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
Black	0.023***	0.023***	-0.059***	-0.080***	-0.080***	-0.147***
	(0.004)	(0.004)	(0.005)	(0.008)	(0.008)	(0.009)
Other ethnicity	0.037***	0.037***	0.002	-0.050***	-0.050***	-0.072***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)
Standardised GCSE English grade	-0.119***	-0.119***	-0.105***	-0.123***	-0.123***	-0.110***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Standardised GCSE Mathematics grade	-0.116***	-0.116***	-0.102***	-0.124***	-0.124***	-0.108***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	406,623	406,623	299,755	345,110	345,110	254,710
Time FE	X	X	X	X	X	X

Notes: In regressions in columns (1)-(3) the sample includes all students. In regressions (4) -(6) the sample is restricted only to students who are not residing in London. The outcome variable is the probability to study in the same geographical region as the one where the student was domiciled at age 16. Regressions (1), (2), (4) and (5) include the treatment dummy as well and cover period 2004/05-2007/08. Regressions (3) and (6) include the individual treatment dummies and cover periods 2004/05-2006/07. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE 3.10: Effect on the probability to live with parents

	All students			Only students not living in London		
	Overall	Year on Year	Treatment Heterogeneity	Overall	Year on Year	Treatment Heterogeneity
	(1)	(2)	(3)	(4)	(5)	(6)
After * Treatment	0.003			0.005		
	(0.007)			(0.007)		
After 2006 * Treatment		0.009			0.010	
		(0.008)			(0.008)	
After 2007 * Treatment		-0.002			-0.001	
		(0.008)			(0.008)	
After 2006 * Treatment(2-20)			0.005			0.006
			(0.008)			(0.008)
After 2006 * Treatment(21-40)			0.007			0.009
			(0.008)			(0.008)
After 2006 * Treatment(41-60)			0.009			0.010
			(0.009)			(0.009)
After 2006 * Treatment(61-80)			0.005			0.009
			(0.009)			(0.010)
After 2006 * Treatment(81-100)			0.013			0.017
			(0.010)			(0.011)
Female	0.034***	0.034***	0.030***	0.033***	0.033***	0.029***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Asian	0.204***	0.204***	0.152***	0.164***	0.164***	0.119***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Black	0.043***	0.043***	-0.033***	0.009	0.009	-0.049***
	(0.004)	(0.004)	(0.005)	(0.007)	(0.007)	(0.008)
Other ethnicity	0.082***	0.082***	0.047***	0.009*	0.009*	-0.011*
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Standardised GCSE English grade	-0.123***	-0.123***	-0.109***	-0.124***	-0.124***	-0.110***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Standardised GCSE Mathematics grade	-0.117***	-0.117***	-0.103***	-0.119***	-0.119***	-0.104***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	406,623	406,623	299,755	345,110	345,110	254,710
Time FE	X	X	X	X	X	X

Notes: In regressions in columns (1)-(6) the sample includes all students. In regressions (7) -(12) the sample is restricted only to students who are not residing in London. The outcome variable is the probability to live at home. Regressions (1), (2), (4) and (5) include the treatment dummy as well and cover period 2004/05-2007/08. Regressions (3) and (6) include the individual treatment dummies and cover periods 2004/05-2006/07. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE 3.11: Effect on the probability to study in a rich area

	All students			Only students not living in London		
	Overall	Year on Year	Treatment Heterogeneity	Overall	Year on Year	Treatment Heterogeneity
	(1)	(2)	(3)	(4)	(5)	(6)
After * Treatment	-0.015			-0.016		
	(0.011)			(0.011)		
After 2006 * Treatment		-0.024*			-0.022	
		(0.013)			(0.014)	
After 2007 * Treatment		-0.006			-0.011	
		(0.013)			(0.014)	
After 2006 * Treatment(2-20)			-0.007			-0.003
			(0.014)			(0.014)
After 2006 * Treatment(21-40)			-0.022			-0.020
			(0.014)			(0.014)
After 2006 * Treatment(41-60)			-0.030**			-0.028**
			(0.014)			(0.014)
After 2006 * Treatment(61-80)			-0.039***			-0.040***
			(0.014)			(0.015)
After 2006 * Treatment(81-100)			-0.041***			-0.051***
			(0.014)			(0.015)
Female	0.007***	0.007***	0.010***	0.005***	0.005***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Asian	0.120***	0.120***	0.099***	-0.041***	-0.041***	-0.042***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
Black	0.266***	0.266***	0.255***	0.016**	0.016**	0.030***
	(0.004)	(0.004)	(0.005)	(0.008)	(0.008)	(0.010)
Other ethnicity	0.188***	0.188***	0.183***	0.053***	0.053***	0.053***
	(0.004)	(0.004)	(0.005)	(0.006)	(0.006)	(0.007)
Standardised GCSE English grade	0.064***	0.064***	0.064***	0.079***	0.079***	0.074***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Standardised GCSE Mathematics grade	0.056***	0.056***	0.057***	0.065***	0.065***	0.059***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observations	406,623	406,623	299,755	345,110	345,110	254,710
Time FE	X	X	X	X	X	X

Notes: In regressions in columns (1)-(6) the sample includes all students. In regressions (7) -(12) the sample is restricted only to students who are not residing in London. The outcome variable is the probability to study in a university located in an affluent area. Regressions (1), (2), (4) and (5) include the treatment dummy as well and cover period 2004/05-2007/08. Regressions (3) and (6) include the individual treatment dummies and cover periods 2004/05-2006/07. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

TABLE 3.12: Robustness checks: various control groups

	All students		Only students not living in London	
	Overall	Treatment Heterogeneity	Overall	Treatment Heterogeneity
	(1)	(2)	(3)	(4)
Panel A: Control 1.5%				
After * Treatment	-0.025		-0.027	
	(0.018)		(0.018)	
After 2006 * Treatment		-0.050**		-0.048**
		(0.022)		(0.022)
After 2007 * Treatment		0.001		-0.005
		(0.022)		(0.022)
Observations	406,623	406,623	345,110	345,110
Panel B: Control 2%				
After * Treatment	-0.026		-0.027*	
	(0.016)		(0.016)	
After 2006 * Treatment		-0.043**		-0.038*
		(0.019)		(0.020)
After 2007 * Treatment		-0.009		-0.016
		(0.019)		(0.020)
Observations	406,623	406,623	345,110	345,110
Panel C: Control 2.5%				
After * Treatment	-0.016		-0.018	
	(0.015)		(0.015)	
After 2006 * Treatment		-0.033*		-0.028
		(0.018)		(0.018)
After 2007 * Treatment		0.002		-0.008
		(0.018)		(0.018)
Observations	406,623	406,623	345,110	345,110
Time FE	X	X	X	X

Notes: In regressions in columns (1)-(2) the sample includes all students. In regressions (3) -(4) the sample is restricted only to students who are not residing in London. The outcome variable is the logarithm of the geographical distance between home and the university attended. In panel A the control group is defined as students in the 1.5% percentile of the IDACI distribution. In panel B the control group is defined as students in the 2% percentile of the IDACI distribution. In panel C the control group is defined as students in the 2.5% percentile of the IDACI distribution. All regressions include the treatment dummy as well and cover period 2004/05-2007/08. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Bibliography

Alecke, B., Burgard, C., and Mitze, T. (2013). The Effect of Tuition Fees on Student Enrollment and Location Choice-Interregional Migration, Border Effects and Gender Differences. Ruhr Economic Papers 0404, Ruhr-Universitat Bochum.

Angrist, J. and Krueger, A. (1992). Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery. *National Bureau of Economic Research, Inc; NBER Working Papers*.

Ashenfelter, O. and Krueger, A. B. (1994). Estimates of the Economic Returns to Schooling from a New Sample of Twins. *American Economic Review, American Economic Association*, 84(5):1157–73.

Azmat, G., Guell, M., and Manning, A. (2006). Gender Gap in Unemployment Rates in OECD countries. *Journal of Labour Economics*, 24(1):1–37.

Azmat, G. and Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics*, 30(C):32–40.

Bassanini, A. and Duval, R. (2006). Employment Patterns in OECD Countries: Reassessing the Role of Policies and Institutions. *OECD Economics Department Working Papers*, (486).

Belot, M. and van Ours, J. (2004). Does the Recent Success of Some OECD Countries in Lowering Their Unemployment Rates Lie in the Clever Design of Their Labor Market Reforms? *Oxford Economic Papers*, 4(56):621–642.

Berger, M. C. (1985). The Effect of Cohort Size on Earnings Growth: A Reexamination of the Evidence. *Journal of Political Economy*, 93(3):561–73.

Bertola, G., Blau, F., and Kahn, L. (2007). Labour Market Institutions and Demographic Employment Patterns. *Journal of Population Economics*, 20(4):833–867.

Betts, J. (1998). Educational Crowding Out: Do Immigrants Affect the Educational Attainment of American Minorities? In *Help or Hindrance? The Economic Implications of Immigration for African-Americans*. Russell Sage Foundation.

BIS (2013). International Education Global Growth and Prosperity: An Accompanying Analytical Narrative. Technical Report BIS/13/1082.

Blau, F. and Kahn, L. (2003). Understanding international Differences in Gender Pay. *Journal of Labour Economics*, 21(1):106–144.

Bloom, D., Freeman, R., and Sanders, K. (1987). Labour Market Consequences of Generational Crowding. *European Journal of Population*, 3(2):131–176.

Bloom, D. and Williamson, J. (1998). Demographic Transitions and Economic Miracles in Emerging Asia. *World Bank Economic Review*, 12(3):419–455.

Borjas, G. (2007). Do foreign students crowd out natives from graduate programs? In Ehrenberg, R. and Stephan, P., editors, *Science and the University*. University of Wisconsin Press, Wisconsin.

Borjas, G., Freeman, R., and Katz, L. (1996). Searching for the Effect of Immigration on the Labor Market. *The American Economic Review*, 86(2):246–251.

Borjas, G. J. (2003). The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *Quarterly Journal of Economics*, 4(118).

Bound, J., Jaeger, D., and Baker, R. (1995). Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak. *Journal of the American Statistical Association*, 90:443–450.

Bound, J. and Turner, S. (2007). How Resources Affect Collegiate Attainment. *Journal of Public Economics*, 91:877–899.

Broecke, S. (2015). University rankings: do they matter in the UK? *Education Economics*, 23(2):137–161.

Brunello, G. (2010). The Effects of Cohort Size on European Earnings. *Journal of Population Economics*, 23(1):273–290.

Campante, L. and Chor, D. (2012). Why was the Arab World Poised for Revolution? Schooling, Economic Opportunities and the Arab Spring. *Journal of Economic Perspectives*, 2:167–188.

Card, D. (1990). The Impact of the Mariel Boatlift on the Miami Labor Market. *Industrial and Labor Relations Review*, 2(43):245–257.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In Christofides, L., Grant, E., and Swidinsky, R., editors, *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp Toronto*. University of Toronto Press, Toronto.

Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Market Impacts of Higher Immigration. *Journal of Labour Economics*, 19(1):22–64.

Card, D. (2005). Is The New Immigration Really So Bad? *Economic Journal*, 115:F300–F323.

Card, D. (2009). Immigration and Inequality. *American Economic Review*, (99):1–21.

Card, D. and Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *Quarterly Journal of Economics*, 116.

Croson, R. and Gneezy, U. (2009). Gender Differences in Prefernces. *Journal of Economic Literature*, 47(2):448–474.

Croxford, L. and Raffe, D. (2014). The Impact of the 2012 Tuition Fee Changes on Student Flows across the UKs Internal Borders. Working paper.

Dearden, L., Fitzsimons, E., and Wyness, G. (2011). The Impact of Tuition Fees and Support on University Participation in the UK. IFS Working Paper W11/17, Institute for Fiscal Studies.

Deming, D. J. and Dynarski, S. (2010). Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor. In *Targeting Investments in Children: Fighting Poverty When Resources are Limited*. University of Chicago Press.

Denzler, S. and Wolter, S. C. (2011). Too Far to Go? Does Distance Determine Study Choices? IZA Discussion Papers 5712, Institute for the Study of Labor (IZA).

Dotterweich, D. and Baryla, E. (2005). Non-resident Tuition and Enrollment in Higher Education: Implications for Tuition Pricing. *Education Economics*, 13(4):375–385.

Dustmann, C., Fabbri, F., and Preston, I. (2005). The Impact of Immigration on the British Labour Market. *The Economic Journal*, 115(507):F324–F341.

Dustmann, C., Fabbri, F., Preston, I., and Wadsworth, J. (2003). Labour Market Performance of Immigrants in the UK Labour Market. *Home Office Online Report*, (05/03).

Dustmann, C., Machin, S., and Schonberg, U. (2010). Ethnicity and Educational Achievement in Compulsory Schooling. *The Economic Journal*, 120(546):F272–F297.

Dwenger, N., Storck, J., and Wrohlich, K. (2012). Do tuition fees affect the mobility of university applicants? Evidence from a natural experiment. *Economics of Education Review*, 31(1):155–167.

Eckel, C. and Grossman, P. (2008). Men, women and risk aversion: experimental evidence. In Plott, C. and Smith, V., editors, *Handbook of Experimental Economics Result, vol.1*, chapter 113. Elsevier, New York.

Elmeskov, J., M. J. and Scarpetta, S. (1998). Key Lessons for Labour Market Reforms: Evidence from OCED Countries' Experiences. *Swedish Economic Policy Review*, 2(5):205–252.

Ermisch, J. (1988). British Labour Market Responses to Age Distribution Changes. In Lee, R., Arthur, W., and Rodgers, G., editors, *Economics of Changing Age Distributions in Developed Countries*. Clarendon Press, Oxford.

Fitzenberger, B. and Kohn, K. (2006). Skill Wage Premia, Employment and Cohort Effects: Are Workers in Germany All of the Same Type? *IZA, Discussion paper(2185)*.

Freeman, R. (1979). The Effect of Demographic Factors on Age-Earning Profiles. *Journal of Human Resources*, 14(3):289–318.

Frenette, M. (2006). Too Far to Go On? Distance to School and University Participation. *Education Economics*, 14(1):31–58.

Garloff, A., Pohl, C., and Schanne, N. (2013). Do Small Labour Market Entry Cohorts Reduce Unemployment? *Demographic Research*, 29:379–406.

Genda, Y., Kondo, A., and Ohta, S. (2010). Long-Term Effecs of a Recession at Labour Market Entry in Japan and the United States. *Journal of Human Resources*, 45(1):59–107.

Gibbons, S., Neumayera, E., and Perkinsa, E. (2015). Student satisfaction, league tables and university applications: Evidence from Britain. *Economics of Education Review*, 48:148 – 164.

Gibbons, S. and Vignoles, A. (2012). Geography, choice and participation in higher education in England. *Regional Science and Urban Economics*, 42:98–113.

Giuliano, P. (2007). Living Arrangements in Western Europe: Does Cultural Origin Matter? *Journal of the European Economic Association*, 5(5):927–952.

Goldin, C. and Katz, L. (2002). The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions. *Journal of Political Economy*, 110(4):730–770.

Gould, E., Lavy, V., and Paserman, D. (2009). Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidece. *The Economic Journal*, 119:1243–1269.

Hoxby, C. M. (1998). Do immigrants crowd disadvantaged american natives out of higher education? In *Help or Hindrance?: The Economic Implications of Immigration for African Americans*. Russell Sage Foundation.

Hubner, M. (2012). Do Tuition Fees Affect Enrollment Behavior? Evidence from a 'Natural Experiment' in Germany. *Economics of Education Review*, 31(6):949 – 960.

Hunt, J. (2012). The Impact of Immigration on the Educational Attainment of Natives. *NBER Working Paper No. 18047*.

Hutchinson, J., Johnes, R., Mao, L., Perera, N., Sellen, P., and Treadaway, M. (2016). Education in England. Annual report 2016, Education Policy Institute.

Jackson, O. (2015). Does Immigration Crowd Natives Into or Out of Higher Education? *Working Paper Federal Reserve Bank of Boston*, (15-18).

Jimeno, J. and Rodriguez-Palenzuela, D. (2002). Youth Unemployment in the OECD: Demographic Shifts, Labour Market Institutions and Macroeconomic Shocks. Working Papers 2002-15, FEDEA.

Kahn, L. B. (2010). The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy. *Labour Economics*, 17(2):303–316.

Kato, T. and Sparber, C. (2013). Quotas and Quality: The Effect of H-1B Visa Restrictions on the Pool of Prospective Undergraduate Students from Abroad. *The Review of Economics and Statistics*, 95(1):109–126.

Korenman, S. and Neumark, D. (2000). Cohort Crowding and Youth Labor Markets: A Cross-National Analysis. NBER Comparative labour market series, NBER.

Kyung, W. (1996). In-Migration of College Students to the State of New York. *The Journal of Higher Education*, 67(3):349–358.

Machin, S. and Murphy, R. (2014). Paying Out and Crowding Out? The Globalisation of Higher Education. CEP Discussion Paper CEPDP1299, Center of Economic Performance.

Manacorda, M., Manning, A., and Wadsworth, J. (2012). The Impact of Immigration on the Structure of Wages: Theory and Evidence from Britain. *Journal of the European Economic Association*, 10(1):120–151.

Manacorda, M. and Moretti, E. (2006). Why do Most Italian Youths Live with Their Parents? Intergenerational Transfers and Household Structure. *Journal of the European Economic Association*, 4(4):800–829.

Manning, A. and Petrongolo, B. (2006). The Part-Time Pay Penalty for Women in Britain. *Economic Journal*, 118(256):F28–52.

Neill, C. (2009). Tuition Fees and the Demand for University Places. *Economics of Education Review*, 28(5):561 – 570.

Newhouse, D. and Wolff, C. (2014). Cohort Size and Youth Employment Outcomes. *IZA Discussion Paper, No. 8197*.

Nickell, S. (1997). Unemployment and Labour Market Rigidities: Europe versus North America. *Journal of Economic Perspectives*, 3(11):55–74.

Nunziata, L. (2002). Unemployment, Labour Market Institutions and Shocks. *Nuffield College Working Papers in Economics*, (W16).

OECD (2011). Education at a Glance 2011. Technical report, OECD.

OECD (2012). The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence. Technical report, OECD.

OECD (2013). Education at a Glance 2013: OECD Indicators. Technical report, OECD Education Database and OECD.

Oreopoulos, P., von Wachter, T., and Heisz, A. (2012). The Short- and Long-Term Career Effects of Graduating in a Recession. *American Economic Journal: Applied Economics*, 4(1):1–29.

Ottaviano, G. I. and Peri, G. (2012). Rethinking the Effects of Immigration on Wages. *Journal of the European Economic Association*, 1(10):152–197.

Parisi, L. (2008). Leaving Home and the Chances of Being Poor: The Case of Young People in Southern European Countries. *Labour*, 22(s1):89–114.

Rutter, J. (2016). Back to Basics: Towards a Successful and Cost-Effective Integration Policy. Report, Institute for Public Policy Research.

Sa, F. (2014). The Effect of Tuition Fees on University Applications and Attendance: Evidence from the UK. IZA Discussion Papers 8364, Institute for the Study of Labor (IZA).

Saleheen, J. and Shadforth, C. (2006). The Economic Characteristics of Immigrtants and Their Impact on Supply. *Bank of England Quarterly Bulettin*, Q4:374–385.

Shimer, R. (2001). The Impact Of Young Workers On The Aggregate Labor Market. *The Quarterly Journal of Economics*, 116(3):969–1007.

Skans, O. (2005). Age Effects in Swedish Local Labour Markets. *Economic Letters*, 86:419–426.

Soo, K. (2013). Does anyone use information from university rankings? *Education Economics*, 21(2):176–190.

Spiess, C. K. and Wrohlich, K. (2010). Does Distance Determine Who Attends a University in Germany? *Economics of Education Review*, 29(3):470 – 479.

Strand, S. (2008). Minority ethnic pupils in the Longitudinal Study of Young People in England: Extension Report on Performance in Public Examinations at Age 16. *DCSF Research Report RR-029*.

Strand, S. (2014). Ethnicity, Gender, Social Class and Achievement Gaps at Age 16: Intersectionality and 'Getting it' for the White Working Class. *Research Papers in Education*, 29(2):131–171.

Tuckman, H. P. (1970). Determinants of College Student Migration. *Southern Economic Journal*, 37(2):184–189.

Welch, F. (1979). Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust. *Journal of Political Economy*, 87(5):S65–S98.

Wright, R. E. (1991). Cohort Size and Earnings in Great Britain. *Journal of Population Economics*, 4(4):295–305.

Appendix A

For all three countries considered, only wage and salary off employees and self-employed workers aged 25 to 60 are included. For the US, the gross weekly wages of full-time workers are calculated by dividing annual wage and salary earnings by the number of weeks worked in the previous year for individuals. In order to calculate the hourly wages, the reported number of hours per week that respondents usually worked if they worked during the previous calendar year is used. The sample includes full-time wage and salary workers who were employed during the previous calendar year prior to the March survey and worked at least thirty-five hours a week. Moreover, people between 15-24 enrolled in full or part time studies are discarded - this will not affect the results, as the sample considered is at least 25 years old. Those with wages below 50 and above 2500 1998 dollars are dropped from the sample.

The French LFS provides information on the previous monthly wages which we multiply by $\frac{12}{52}$ in order to convert them to weekly wages. All wages are recalculated in 1998 Euros, using the CPI provided by the French National Institute of Statistics and the pegged value. It is also worth mentioning that in 1999, the French Franc was partially replaced by the Euro, and in 2002, the Euro became the official currency of France. The French Franc was pegged to Euro at the value of 6.55957 Francs. The sample includes people working full-time in their main job the previous month and it excludes those enrolled in studies. Moreover, all weekly wages below 50 and above 2000 1998 euros are dropped. Wages are for those fulltime employed and with nonzero earnings.

For the UK data, weekly gross wages, identified as the weekly earnings in the main job in the reference week are available from 1993 onwards. The values are calculated in 1998 GB, using the CPI from ONS. Only people in full-time employment during the reference period are kept and those in employment who are still in education (whether full or part time). Weekly wages less than 50 and higher than 2000 in 1998 GB pounds are discarded. In order to keep the analysis consistent, we express all wages in the analysis in PPP 1998 US dollars, using the PPP calculated by the World Bank.

Appendix B

GCSE

For the period under analysis, the grading system of the GCSEs changed. Based on the information provided by Ofsted and Ofqual, the following scales were used in the calculation of the grades obtained in the GCSE in English and in Maths:

Table B1: Grading system GCSEs

Panel A:Single Awards															
Grade	A*	A	B	C	D	E	F	G							
Old points(before 2004)	8	7	6	5	4	3	2	1							
New points(2004 onwards)	58	52	46	40	34	28	22	16							
Panel B: Double Awards															
Grade	A*A*	A*A	AA	AB	BB	BC	CC	CD	DD	DE	EE	EF	FF	FG	GG
New points (2008 onwards)	58	55	52	49	46	43	40	37	34	31	28	25	22	19	16

Notes: Double Award GCSE subjects are certificated on a fifteen-point scale for the first time in the June 2008 examination. For the Double Awards, the grade is recorded twice on the certificate to indicate that the results in these specifications have the same status as GCSE grades in two other single-certificate subjects. Source Ofsted, Ofqual

Higher Education Institutions

In 1998/99 there were 170 state funded universities that reported their data on student numbers to HESA. Until 2011/12, 17 of them merged with other existing universities, while 6 merged to form 3 new universities and 5 new universities opened from splitting existing ones. These universities that merged or split are treated as one, which makes an balanced list of 150 universities. The 11 new universities which emerged during this period are not considered, nor is the university that closed.

Out of these 150, only 139 have full-time undergraduate students (Open University, Cranfield University, Royal College of Art, Royal College of Nursing,Bishop Grosseteste University, Liverpool Hope University, Birkbeck College, Institute of Education, London Business School, London School of Hygiene and Tropical Medicine).

Field of Study

In the HESA data there are 20 major field of study pursued at higher education level, but given that the number of students from specific countries in each field of study group in each year is very small, we group the fields of study in 5 groups as below:

Table B2: Coding of field of study

JACS Subject Groups	5 Subject Groups
Medicine and Dentistry	Medicine, Dentistry and Allied Subjects
Other Medical Subjects	Medicine, Dentistry and Allied Subjects
Biological Sciences	Medicine, Dentistry and Allied Subjects
Veterinary Sciences and Agriculture	Medicine, Dentistry and Allied Subjects
Physical Sciences	STEM
Maths and Computer Sciences	STEM
Engineering	STEM
Technology	STEM
Architecture,Building and Planning	STEM
Social Sciences	Social Sciences
Law	Social Sciences
Business and Administration	Social Sciences
Mass Communication and Documentation	Languages and History
Linguistics and Classics	Languages and History
European Languages	Languages and History
Modern Languages	Languages and History
History and Philosophical Studies	Languages and History
Creative Arts and Design	Arts, Education, Other
Education	Arts, Education, Other
Combined	Arts, Education, Other

Undergraduate degree definition

The undergraduate students who represent the student population considered in this analysis are formed or two categories of students: first degree and other undergraduate degree. According to HESA, the First degree includes first degrees with or without eligibility to register to practice with a Health or Social Care or Veterinary statutory regulatory body, first degrees with qualified teacher status (QTS)/registration with the General Teaching Council (GTC), enhanced first degrees, first degrees obtained concurrently with a diploma and intercalated first degrees. Other undergraduate includes qualification aims below degree level such as Foundation Degrees, diplomas in HE with eligibility to register to practice with a Health or Social Care regulatory body, Higher National Diploma (HND), Higher National Certificate (HNC), Diploma of Higher Education (DipHE),Certificate of Higher Education (CertHE), foundation courses at HE level, NVQ/SVQ levels 4 and 5, post-degree diplomas and certificates at undergraduate level, professional qualifications at undergraduate level, other undergraduate diplomas and certificates including post registration health and social care courses, other formal HE qualifications of less than degree standard, institutional undergraduate credit and no formal undergraduate qualifications. The coding also accounts for the mapping between the old and the new codes which was introduced in 2007/08.[1]

Domicile

We defined student's country of origin using the domicile reported in the HESA data. We used the mapping of the domicile codes provided by HESA in order to account for the changes in the coding from 2007/08 onwards.[2]

[1]Source: HESA undergraduate degree mapping. Website https://www.hesa.ac.uk/data-and-analysis/performance-indicators/definitions#level-study-applicable-all-tables

[2]Source: HESA Domicile mapping. Website: https://www.hesa.ac.uk/support/definitions/students#domicile